



# Consistent estimation of a mean planar curve modulo similarities

Jérémie Bigot, Benjamin Charlier

## ► To cite this version:

Jérémie Bigot, Benjamin Charlier. Consistent estimation of a mean planar curve modulo similarities. 2012. hal-00634899v2

**HAL Id: hal-00634899**

**<https://hal.science/hal-00634899v2>**

Submitted on 31 Oct 2012

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Consistent estimation of a mean planar curve modulo similarities

J  r  mie Bigot\*, DMIA/ISAE - Universit   de Toulouse  
Benjamin Charlier†, IMT/CNRS (UMR 5219) - Universit   de Toulouse

October 31, 2012

## Abstract

We consider the problem of estimating a mean planar curve from a set of  $J$  random planar curves observed on a  $k$ -points deterministic design. We study the consistency of a smoothed Procrustean mean curve when the observations obey a deformable model including some nuisance parameters such as random translations, rotations and scaling. The main contribution of the paper is to analyze the influence of the dimension  $k$  of the data and of the number  $J$  of observed configurations on the convergence of the smoothed Procrustean estimator to the mean curve of the model. Some numerical experiments illustrate these results.

*Keywords* : Shape spaces; Perturbation models; Consistency; Procrustes Means; Nonparametric inference; High-dimensional data; Linear smoothing.

*AMS classifications*: Primary 62G08; secondary 62H11.

## Acknowledgements

The authors acknowledge the support of the French Agence Nationale de la Recherche (ANR) under reference ANR-JCJC-SIMI1 DEMOS.

## 1 Introduction

In this paper, we are interested in the statistical treatment of random planar curves observed through rigid deformations of the plane. In many fields of interest, such data appear as, for instance, contours extracted from digital images or level sets of a real function defined on the plane. In handwriting recognition problems one typically compares curves that describe letters, digits or signatures and the acquisition process often create some ambiguity of location, size and orientation. Our aim is then to study an estimation procedure for a mean curve from a sample of  $J$  noisy and discretized planar curves observed through translation, rotations and scaling. The group generated by this set of transformations is usually called *similarity group* of the plane and in the sequel, an element of this group will be called a *deformation*.

### 1.1 A deformable model coming from statistical shape analysis

**Deformable model** In many practical cases of interest, data are collected through a computer device such as a digital camera or an image scanner. In order to take this into account in our model, we assume that observations at hand are discretized versions of continuous planar curves. Each observation  $\mathbf{Y}$  is

---

\*Institut Sup  rieur de l'A  ronautique et de l'Espace, D  partement Math  matiques, Informatique, Automatique, 10 Avenue   douard-Belin, BP 54032-31055, Toulouse CEDEX 4, France. Email: jeremie.bigot@isae.fr

†Institut de Math  matiques de Toulouse, Universit   de Toulouse et CNRS (UMR 5219), 31062 Toulouse, Cedex 9, France. Email: benjamin.charlier@math.univ-toulouse.fr

then given by a set of  $k$  points of the plane called a *configuration*. It can be written as a  $k \times 2$  real matrix

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_k \end{pmatrix} \in \mathbb{R}^{k \times 2},$$

where  $Y_\ell = (Y_\ell^{(1)}, Y_\ell^{(2)}) \in \mathbb{R}^2$  with  $\ell = 1, \dots, k$ . A degenerated configuration is a configuration composed of  $k$  times the same point  $b = (b^{(1)}, b^{(2)}) \in \mathbb{R}^2$ . In tensorial notation, such a configuration is written  $\mathbf{1}_k \otimes b \in \mathbb{R}^{k \times 2}$  where  $\mathbf{1}_k$  denotes the (column) vector of  $\mathbb{R}^k$  with all entries equal to one and  $\otimes$  denotes the tensor product. From now on, we assume that the observations satisfy the following functional regression model for  $j = 1, \dots, J$ ,

$$\mathbf{Y}_j = e^{a_j^*}(\mathbf{f} + \boldsymbol{\zeta}_j)R_{\alpha_j^*} + \mathbf{1}_k \otimes b_j^*, \quad \text{with } R_{\alpha_j^*} = \begin{pmatrix} \cos(\alpha_j^*) & -\sin(\alpha_j^*) \\ \sin(\alpha_j^*) & \cos(\alpha_j^*) \end{pmatrix}, \quad (1.1)$$

where the unknown mean pattern  $\mathbf{f} \in \mathbb{R}^{k \times 2}$  has been obtained by sampling on an equi-spaced design a planar curve  $f = (f^{(1)}, f^{(2)}) : [0, 1] \rightarrow \mathbb{R}^2$  satisfying  $f(0) = f(1)$ . It means that we have

$$\mathbf{f} = \left(f\left(\frac{\ell}{k}\right)\right)_{\ell=1}^k := \left(f^{(1)}\left(\frac{\ell}{k}\right), f^{(2)}\left(\frac{\ell}{k}\right)\right)_{\ell=1}^k \in \mathbb{R}^{k \times 2}.$$

The error terms  $\boldsymbol{\zeta}_j \in \mathbb{R}^{k \times 2}$ ,  $j = 1, \dots, J$  are independent copies of a random perturbation  $\boldsymbol{\zeta}$  in  $\mathbb{R}^{k \times 2}$  with zero expectation. For  $j = 1, \dots, J$ , the scaling, rotation and translation parameters  $(a_j^*, \alpha_j^*, b_j^*) \in \mathbb{R} \times [-\pi, \pi] \times \mathbb{R}^2$  are independent and identically distributed (i.i.d) random variables independent of the random perturbations  $\boldsymbol{\zeta}_j$ .

Model (1.1) is a deformable model in the sens of [Bigot & Charlier, 2011] and our aim is to estimate the underlying mean curve  $f : [0, 1] \rightarrow \mathbb{R}^2$  from  $\mathbf{Y}_1, \dots, \mathbf{Y}_J \in \mathbb{R}^{k \times 2}$ . More precisely, we study the influence on the estimation procedure of the number  $J$  of configurations at hand and the number  $k$  of discretization points composing the configurations  $\mathbf{Y}_j$ 's. Note that the model (1.1) has been introduced by Goodall in [Goodall, 1991] but with a fixed number  $k$  of points in each configuration and non-random nuisance parameters  $(a_j^*, \alpha_j^*, b_j^*)$ 's. This framework has been highly popular in the statistical shape community, see [Kent & Mardia, 1997, Le, 1998, Lele, 1993]. Note that the configuration  $\mathbf{f}$  was called a population mean in [Goodall, 1991] or a perturbation mean in [Huckemann, 2011].

**Shape of a configuration.** Since the seminal work of Kendall [Kendall, 1984], one considers that the shape of a configuration  $\mathbf{Y} \in \mathbb{R}^{k \times 2}$  is “what remains when translations, rotations and scaling are filtered out”. More precisely, two configurations  $\mathbf{Y}_1, \mathbf{Y}_2 \in \mathbb{R}^{k \times 2}$  are said to have the same shape if there exists a vector  $(a, \alpha, b) \in \mathbb{R} \times [-\pi, \pi] \times \mathbb{R}^2$  such that

$$\mathbf{Y}_2 = e^a \mathbf{Y}_1 R_\alpha + \mathbf{1}_k \otimes b, \quad \text{with } R_\alpha = \begin{pmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{pmatrix}, \quad (1.2)$$

The Kendall's Shape Space is the quotient space modulo this equivalent relation. It is usually denoted by  $\Sigma_2^k$  and defined as the set of normalized (*i.e.* centered and scaled to size one) configurations quotiented by the rotation of the plane, see part 2.2 for further details.

The above definition can be trivially extended to the case of planar curves by replacing the configuration  $\mathbf{f} \in \mathbb{R}^{k \times 2}$  by the planar curve  $f : [0, 1] \rightarrow \mathbb{R}^2$ . Thence, we call *shape* of  $f$  the set of planar curves that can be written  $e^a f R_\alpha + \mathbf{1} \otimes b$  for some  $(a, \alpha, b) \in \mathbb{R} \times [-\pi, \pi] \times \mathbb{R}^2$ .

## 1.2 Estimation of the mean curve and of the shape of the mean curve

**Consistency in the Shape Space.** In shape analysis, an important issue is the computation of a sample mean shape from a set of  $J$  random planar configurations  $\mathbf{Y}_1, \dots, \mathbf{Y}_J \in \mathbb{R}^{k \times 2}$  satisfying

model (1.1) and the study of its consistency as the number of samples  $J$  goes to infinity ( $k$  remaining fixed). According to Goodall [Goodall, 1991] a sample mean pattern  $\hat{\mathbf{f}} \in \mathbb{R}^{k \times 2}$  computed from  $\mathbf{Y}_1, \dots, \mathbf{Y}_J$  is said to be consistent if, as  $J \rightarrow \infty$ , it has asymptotically the same shape than the mean pattern  $\mathbf{f}$ . In this framework, the deformation parameters are considered as *nuisance parameters* that contain no informations. That is why the data are first normalized (*i.e.* centered and scale to unit size) without changes in the statistical analysis. The study of consistent procedures to estimate the shape of the mean pattern  $\mathbf{f}$  using this approach has been considered by various authors [Kent & Mardia, 1997, Le, 1998, Lele, 1993, Huckemann, 2011]. In this setting, sample mean patterns obtained by a Procrustes procedure have received a special attention. In particular, it is shown in [Kent & Mardia, 1997, Le, 1998] that, in the very specific case of isotropic perturbations  $\boldsymbol{\zeta}$ , the so-called full and partial Procrustes sample means are consistent estimators of the shape of  $\mathbf{f}$ . Nevertheless, these estimators can be inconsistent for non-isotropic perturbations. Therefore, it is generally the belief that consistent statistical inference based on Procrustes analysis is restricted to very limited assumptions on the distribution of the data, see also [Dryden & Mardia, 1998, Kendall et al., 1999, Huckemann, 2011] for further discussions.

**Estimation of the mean curve.** Our approach differ from the one developed in Shape analysis in two main aspects. First, we assume that the unknown mean pattern  $\mathbf{f} \in \mathbb{R}^{k \times 2}$  in model (1.1) has been obtained by sampling a planar curve  $f : [0, 1] \rightarrow \mathbb{R}^2$  on an equi-spaced design. This allows us to study the influence of the number  $k$  of points composing each configuration on the estimation of  $\mathbf{f}$ . Note that if  $f$  is sufficiently regular and  $k$  is large, to estimate  $\mathbf{f} \in \mathbb{R}^{k \times 2}$  (*i.e.* the values of  $f$  on the design) roughly amounts to estimate  $f$ . The second difference concerns the randomness of the deformation parameters  $(a_j^*, \alpha_j^*, b_j^*) \in \mathbb{R} \times [-\pi, \pi] \times \mathbb{R}^2$ ,  $j = 1, \dots, J$ , that we want to estimate rather than to consider as nuisance parameters. Indeed, the values of the deformations parameters can be informative in some cases : assume that the size of the data are very similar except for one. This difference in size may be due to a (possibly) relevant factor and if the data are normalized as in Shape analysis this information is lost.

Under a suitable smoothness assumption on  $f$ , we are able to estimate consistently, with an asymptotic in  $k$  only, the curve  $f_j := e^{a_j^*} f R_{\alpha_j^*} + \mathbf{1} \otimes b_j^*$  from  $\mathbf{Y}_j$  for each  $j = 1, \dots, J$ . Note that, by definition, the  $f_j$ 's have the same shape as  $f$ . When  $J$  is fixed, this smoothing step allows us to estimate consistently the deformations parameters  $(a_j^*, \alpha_j^*, b_j^*)_{j=1}^J$  with a Procrustes matching step. Now, if we assume that the deformations parameters  $(a_j^*, \alpha_j^*, b_j^*)$  are centered random variables, we show that it is possible to recover the *true* mean shape  $f$  when  $\min\{J, k\} \rightarrow \infty$ . These results are rigorously stated in the next section, see Theorem 1.1 below.

### 1.3 Main contribution

Our estimating procedure is composed of two steps. First, we perform a dimension reduction step by projecting the data into a low-dimensional space of  $\mathbb{R}^{k \times 2}$  to eliminate the influence of the random perturbations  $\boldsymbol{\zeta}_j$ . Then, in a second step, we apply Procrustes analysis in this low-dimensional space to obtain a consistent estimator of  $\mathbf{f} \in \mathbb{R}^{k \times 2}$ .

The reduction dimension step is based on an appropriate smoothness assumptions on  $f$ . Let  $L > 0$ ,  $s > 0$  and define the Sobolev ball of radius  $L$  and degree  $s$  as

$$H_s(L) = \left\{ f = (f^{(1)}, f^{(2)}) \in L^2([0, 1], \mathbb{R}^2), \sum_{m \in \mathbb{Z}} (1 + |m|^2)^s (|c_m(f^{(1)})|^2 + |c_m(f^{(2)})|^2) < L \right\} \quad (1.3)$$

where  $c_m(f) = (c_m(f^{(1)}), c_m(f^{(2)})) = \left( \int_0^1 f^{(1)}(t) e^{-i2\pi m t} dt, \int_0^1 f^{(2)}(t) e^{-i2\pi m t} dt \right) \in \mathbb{C}^2$  is the  $m$ -th Fourier coefficient of  $f = (f^{(1)}, f^{(2)}) \in L^2([0, 1], \mathbb{R}^2)$ , for  $m \in \mathbb{Z}$ .

**Assumption 1.** *The curve  $f : [0, 1] \rightarrow \mathbb{R}^2$  is closed, i.e.  $f(0) = f(1) \in \mathbb{R}^2$ , and belongs to  $H_s(L)$  for some  $L > 0$  and  $s > \frac{3}{2}$ . Moreover, the  $k \times 2$  matrix  $\mathbf{f} = (f(\frac{\ell}{k}))_{\ell=1}^k$  is of rank two, i.e. the configuration  $\mathbf{f}$  is not degenerated.*

Assumption 1 implies that  $f$  is not reduced to a point, continuously differentiable and is equal to its Fourier series. We introduce the following  $k \times k$  matrix

$$\mathbf{A}^\lambda = \left( \frac{1}{k} \sum_{0 \leq |m| \leq \lambda} e^{i2\pi m \frac{\ell - \ell'}{k}} \right)_{\ell, \ell'=1}^k. \quad (1.4)$$

The matrix  $\mathbf{A}^\lambda$  is the smoothing matrix corresponding to a discrete Fourier low pass filter with frequency cutoff  $\lambda \in \mathbb{N}$ . It is a projection matrix in a sub-space  $\mathbf{V}^\lambda$  of  $\mathbb{R}^k$  of dimension  $2\lambda + 1$ . Then, we project the data on  $\mathbf{V}^\lambda \times \mathbf{V}^\lambda \subset \mathbb{R}^{k \times 2}$ , and we estimate the scaling, rotation and translation parameters in model (1.1) using M-estimation as follows: denote the scaling parameters by  $\mathbf{a} = (a_1, \dots, a_J) \in \mathbb{R}^J$ , the rotation parameters by  $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_J) \in \mathbb{R}^J$  and the translation parameters by  $\mathbf{b} = (b_1, \dots, b_J) \in \mathbb{R}^{2J}$ , and introduce the functional,

$$M^\lambda(\mathbf{a}, \boldsymbol{\alpha}, \mathbf{b}) = \frac{1}{Jk} \sum_{j=1}^J \left\| e^{-a_j} \mathbf{A}^\lambda(\mathbf{Y}_j - \mathbf{1}_k \otimes b_j) R_{-\alpha_j} - \frac{1}{J} \sum_{j'=1}^J e^{-a_{j'}} \mathbf{A}^\lambda(\mathbf{Y}_{j'} - \mathbf{1}_k \otimes b_{j'}) R_{-\alpha_{j'}} \right\|_{\mathbb{R}^{k \times 2}}^2, \quad (1.5)$$

where  $\|\cdot\|_{\mathbb{R}^{k \times 2}}$  is the standard Euclidean norm in  $\mathbb{R}^{k \times 2}$ . An M-estimator of

$$(\mathbf{a}^*, \boldsymbol{\alpha}^*, \mathbf{b}^*) = (a_1^*, \dots, a_J^*, \alpha_1^*, \dots, \alpha_J^*, b_1^*, \dots, b_J^*) \in \mathbb{R}^J \times [-\pi, \pi]^J \times \mathbb{R}^{2J}$$

is given by

$$(\hat{\mathbf{a}}^\lambda, \hat{\boldsymbol{\alpha}}^\lambda, \hat{\mathbf{b}}^\lambda) \in \underset{(\mathbf{a}, \boldsymbol{\alpha}, \mathbf{b}) \in \Theta_0}{\operatorname{argmin}} M^\lambda(\mathbf{a}, \boldsymbol{\alpha}, \mathbf{b}), \quad (1.6)$$

where  $(\hat{\mathbf{a}}^\lambda, \hat{\boldsymbol{\alpha}}^\lambda, \hat{\mathbf{b}}^\lambda) = (\hat{a}_1^\lambda, \dots, \hat{a}_J^\lambda, \hat{\alpha}_1^\lambda, \dots, \hat{\alpha}_J^\lambda, \hat{b}_1^\lambda, \dots, \hat{b}_J^\lambda) \in \mathbb{R}^J \times [-\pi, \pi]^J \times \mathbb{R}^{2J}$  and

$$\Theta_0 = \left\{ (\mathbf{a}, \boldsymbol{\alpha}, \mathbf{b}) \in [-A, A]^J \times [-\mathcal{A}, \mathcal{A}]^J \times \mathbb{R}^{2J} : \sum_{j=1}^J a_j = 0, \sum_{j=1}^J \alpha_j = 0 \text{ and } \sum_{j=1}^J b_j = 0 \right\}, \quad (1.7)$$

with  $A, \mathcal{A} > 0$  being parameters whose values will be discussed below.

Finally, the mean pattern  $\mathbf{f}$  is estimated by the following smoothed Procrustes mean

$$\hat{\mathbf{f}}^\lambda = \frac{1}{J} \sum_{j=1}^J e^{-\hat{a}_j^\lambda} \left( \mathbf{A}^\lambda \mathbf{Y}_j - \mathbf{1}_k \otimes \hat{b}_j^\lambda \right) R_{-\hat{\alpha}_j^\lambda}. \quad (1.8)$$

To analyze the convergence of the estimator  $\hat{\mathbf{f}}^\lambda$  to the mean pattern  $\mathbf{f}$ , let us introduce some regularity conditions on the covariance structure of the random variable  $\boldsymbol{\zeta}$  in  $\mathbb{R}^{k \times 2}$ . Let  $\tilde{\boldsymbol{\zeta}} = (\boldsymbol{\zeta}^{(1)}, \boldsymbol{\zeta}^{(2)}) = (\zeta_1^{(1)}, \dots, \zeta_k^{(1)}, \zeta_1^{(2)}, \dots, \zeta_k^{(2)}) \in \mathbb{R}^{2k}$  be the vectorized version of  $\boldsymbol{\zeta} = (\zeta_\ell^{(1)}, \zeta_\ell^{(2)})_{\ell=1}^k \in \mathbb{R}^{k \times 2}$ .

**Assumption 2.** *The random variable  $\tilde{\boldsymbol{\zeta}}$  is a centered Gaussian vector in  $\mathbb{R}^{2k}$  with covariance matrix  $\boldsymbol{\Sigma}_k \in \mathbb{R}^{2k \times 2k}$ . Let  $\gamma_{\max}(k)$  be the largest eigenvalue of  $\boldsymbol{\Sigma}_k$ . Then,*

$$\lim_{k \rightarrow \infty} \gamma_{\max}(k) k^{-\frac{2s}{2s+1}} = 0,$$

where  $s$  is the smoothness parameter defined in Assumption 1.

For example, the isotropic Gaussian error model corresponds to  $\boldsymbol{\Sigma}_k = Id_k$  and  $\gamma_{\max}(k) = 1$  satisfies Assumption 2. If there exists correlations terms (i.e. non-zero off-diagonal entries in  $\boldsymbol{\Sigma}_k$ ), then the level of the perturbation  $\tilde{\boldsymbol{\zeta}}$  has to be sufficiently small. A simple model is the case where  $\boldsymbol{\Sigma}_k = [S(|\ell - \ell'|)]_{\ell, \ell'=1}^k$  for some function  $S : \mathbb{R} \rightarrow \mathbb{R}$  satisfying  $\int_{\mathbb{R}} |S(t)| dt < +\infty$  implying that  $\gamma_{\max}(k) \leq \sum_{\ell \in \mathbb{Z}} |S(\ell)|$  (see Lemma 4.11 in [Gray, 2006]) and thus Assumption 2 is again satisfied. The following theorem is the main result of the paper.

**Theorem 1.1.** Consider model (1.1) and suppose that Assumptions 1 and 2 hold. Suppose also that the random variables  $(\mathbf{a}^*, \boldsymbol{\alpha}^*, \mathbf{b}^*)$  are bounded and belong to  $[-\frac{A}{2}, \frac{A}{2}]^J \times [-\frac{A}{2}, \frac{A}{2}]^J \times [-B, B]^{2J}$  for some  $0 < A, B$  and  $0 < \mathcal{A} < \frac{\pi}{4}$  and that  $\lambda(k) = k^{\frac{1}{2s+1}}$ .

- For any  $J \geq 2$  there exists  $\mathbf{f}_{\Theta_0} = e^{a_0} \mathbf{f} R_{\alpha_0} + \mathbf{1}_k \otimes b_0$  for some  $(a_0, \alpha_0, b_0) \in \mathbb{R} \times [-\pi, \pi[ \times \mathbb{R}^2$  and a function  $V_1(k, x)$  such that for any  $x > 0$ ,

$$\mathbb{P} \left( \frac{1}{k} \|\hat{\mathbf{f}}^\lambda - \mathbf{f}_{\Theta_0}\|_{\mathbb{R}^{k \times 2}}^2 \geq V_1(k, x) \right) \leq e^{-x}, \quad (1.9)$$

with  $V_1(k, x) \rightarrow 0$  when  $k \rightarrow \infty$  and  $x$  remains fixed.

- Suppose, in addition, that the random variables  $(\mathbf{a}^*, \boldsymbol{\alpha}^*, \mathbf{b}^*)$  have zero expectation in  $[-\frac{A}{2}, \frac{A}{2}]^J \times [-\frac{A}{2}, \frac{A}{2}]^J \times [-B, B]^{2J}$  with  $A, \mathcal{A} < 0.1$ . Then, there exists a function  $V_2(J, k, x)$  such that for any  $x > 0$ ,

$$\mathbb{P} \left( \frac{1}{k} \|\hat{\mathbf{f}}^\lambda - \mathbf{f}\|_{\mathbb{R}^{k \times 2}}^2 \geq V_2(J, k, x) \right) \leq e^{-x}, \quad (1.10)$$

where  $V_2(J, k, x) \rightarrow 0$  when  $\min\{J, k\} \rightarrow \infty$  and  $x$  remains fixed.

Statement (1.9) means that, under mild assumptions on the covariance structure of the error terms  $\zeta_j$ , it is possible to consistently estimate the shape of the mean curve  $f$  when the number of observations  $J$  is fixed and the number  $k$  of discretization points increases. Note that  $(a_0, \alpha_0, b_0)$  depends on  $J$  and is given by formula (3.3) in Section 3. The function  $V_1(k, x)$  is explicitly given in Section 5.2. To obtain statement (1.10), we assume the condition  $A, \mathcal{A} < 0.1$  which means that the random scaling and rotations in model (1.1) are not too large. Also, it is assumed that random scaling, rotations and translations have zero expectation, meaning that the deformations parameters in model (1.1) are centered around the identity. Then, under such assumptions, statement (1.10) shows that one can consistently estimate the true mean curve  $f$  when both the sample size  $J$  and the number of landmarks  $k$  go to infinity. Again, the function  $V_2(J, k, x)$  is explicitly given in Section 5.2. These results are consistent with those obtained in [Bigot & Charlier, 2011], where we have studied the consistency of Fréchet means in deformable models for signal and image processing.

## 1.4 Organization of the paper

In Section 2, we recall some properties on the similarity group of the plane, and we describe its action on the mean pattern  $\mathbf{f}$ . Then, we discuss General Procrustes Analysis (GPA) and we compare it to our approach. In Section 3 we discuss some identifiability issues in model (1.1). The estimating procedure is described in detail in Section 4. Consistency results are given in Section 5. Some experiments in Section 6 illustrate the numerical performances of our approach. All the proofs are gathered in a technical appendix.

## 2 Group structure and Generalized Procrustes Analysis

### 2.1 The similarity group

**Group action** First let us introduce some notations and definitions that will be useful throughout the paper. The similarity group of the plane is the group  $(\mathcal{G}, \cdot)$  generated by isotropic scaling, rotations and translations. The identity element in  $\mathcal{G}$  is denoted  $e$  and the inverse of  $g \in \mathcal{G}$  is denoted by  $g^{-1}$ . We parametrize the group  $\mathcal{G}$  by a scaling parameter  $a \in \mathbb{R}$ , an angle  $\alpha \in [-\pi, \pi[$  and a translation  $b \in \mathbb{R}^2$ , and we make no difference between  $g \in \mathcal{G}$  and its parametrization  $(a, \alpha, b) \in \mathbb{R} \times [-\pi, \pi[ \times \mathbb{R}^2$ .

For all  $g_1 = (a_1, \alpha_1, b_1), g_2 = (a_2, \alpha_2, b_2) \in \mathbb{R} \times [-\pi, \pi] \times \mathbb{R}^2$  we have

$$\begin{aligned} g_1 \cdot g_2 &= (a_1, \alpha_1, b_1) \cdot (a_2, \alpha_2, b_2) = (a_1 + a_2, \alpha_1 + \alpha_2, e^{a_1} b_2 R_{\alpha_1} + b_1), \\ g_1^{-1} &= (a_1, \alpha_1, b_1)^{-1} = (-a_1, -\alpha_1, -e^{-a_1} b_1 R_{-\alpha_1}), \\ e &= (0, 0, 0). \end{aligned} \quad (2.1)$$

The action of  $\mathcal{G}$  onto  $\mathbb{R}^{k \times 2}$  is given by the mapping  $(g, \mathbf{x}) \mapsto g \cdot \mathbf{x} := e^a \mathbf{x} R_\alpha + \mathbf{1}_k \otimes b$ , for  $g = (a, \alpha, b) \in \mathcal{G}$  and  $\mathbf{x} \in \mathbb{R}^{k \times 2}$ . Note that we use the same symbol “.” for the composition law of  $\mathcal{G}$  and its action on  $\mathbb{R}^{k \times 2}$ . This action can also be defined on  $L^2([0, 1], \mathbb{R}^2)$  by replacing  $\mathbf{f}$  by  $f \in L^2([0, 1], \mathbb{R}^2)$ . Coming back to the final dimensional case, let

$$\mathbf{1}_{k \times 2} = \mathbf{1}_k \otimes \mathbb{R}^2$$

be the two dimensional linear subspace of  $\mathbb{R}^{k \times 2}$  consisting of degenerated configurations, *i.e.* configurations composed of  $k$  times the same landmarks. The orthogonal subspace  $\mathbf{1}_{k \times 2}^\perp$  is the set of centered configurations. We have the orthogonal decomposition  $\mathbb{R}^{k \times 2} = \mathbf{1}_{k \times 2}^\perp \oplus \mathbf{1}_{k \times 2}$ , and for any configuration  $\mathbf{x} \in \mathbb{R}^{k \times 2}$  we write  $\mathbf{x} = \mathbf{x}_0 + \bar{\mathbf{x}} \in \mathbf{1}_{k \times 2}^\perp \oplus \mathbf{1}_{k \times 2}$ . We call  $\mathbf{x}_0$  the centered configuration of  $\mathbf{x}$  and  $\bar{\mathbf{x}} = \mathbf{1}_k \otimes \left( \frac{1}{k} \sum_{\ell=1}^k x_\ell^{(1)}, \frac{1}{k} \sum_{\ell=1}^k x_\ell^{(2)} \right)$  the degenerated configuration associated to  $\mathbf{x}$ , see Figure 1 for an illustration.

**Orbit, stabilizer and section** Given a configuration  $\mathbf{x}$  in  $\mathbb{R}^{k \times 2}$ , the *orbit* of  $\mathbf{x}$  is defined as the set

$$\mathcal{G} \cdot \mathbf{x} := \{g \cdot \mathbf{x}, g \in \mathcal{G}\} \subset \mathbb{R}^{k \times 2}.$$

This set is also called the shape of  $\mathbf{x}$ . The orbit of any degenerated configuration is the entire subspace  $\mathbf{1}_{k \times 2}$ . Note also that the linear subspace  $\mathbf{1}_{k \times 2}$  is stable by the action of  $\mathcal{G}$ , and that the action of  $\mathcal{G}$  on  $\mathbf{1}_{k \times 2}$  is not free, meaning that for any  $\bar{\mathbf{x}} \in \mathbf{1}_{k \times 2}$  the equality  $g_1 \cdot \bar{\mathbf{x}} = g_2 \cdot \bar{\mathbf{x}}$  does not imply that  $g_1 = g_2$ . Now, if  $\mathbf{x} \in \mathbb{R}^{k \times 2} \setminus \mathbf{1}_{k \times 2}$  is a non-degenerated configuration of  $k$  landmarks, its orbit  $\mathcal{G} \cdot \mathbf{x}$  is a sub-manifold of  $\mathbb{R}^{k \times 2} \setminus \mathbf{1}_{k \times 2}$  of dimension  $\dim(\mathcal{G}) = 4$ .

Given a configuration  $\mathbf{x} \in \mathbb{R}^{k \times 2}$ , the *stabilizer*  $I(\mathbf{x})$  is the closed subgroup of  $\mathcal{G}$  which leaves  $\mathbf{x}$  invariant, namely

$$I(\mathbf{x}) = \{g \in \mathcal{G} : g \cdot \mathbf{x} = \mathbf{x}\}.$$

If  $\bar{\mathbf{x}} = \mathbf{1}_k \otimes (\bar{x}^{(1)}, \bar{x}^{(2)}) \in \mathbf{1}_{k \times 2}$  is a degenerated configuration, its stabilizer is non trivial and is equal to  $I(\bar{\mathbf{x}}) = \{(a, \alpha, (\bar{x}^{(1)}, \bar{x}^{(2)}) - e^a (\bar{x}^{(1)}, \bar{x}^{(2)}) R_\alpha), a \in \mathbb{R}, \alpha \in [-\pi, \pi]\}$ . If  $\mathbf{x} \in \mathbb{R}^{k \times 2} \setminus \mathbf{1}_{k \times 2}$  is a non-degenerated configuration, its stabilizer  $I(\mathbf{x})$  is trivial, *i.e.* is reduced to the identity  $\{e\}$ . The action of  $\mathcal{G}$  is said free if the stabilizer of any point is reduced to the identity. Hence, the action of  $\mathcal{G}$  is free on the set of non-degenerated configurations of  $k$ -ads in  $\mathbb{R}^2$ .

A *section* of the orbits of  $\mathcal{G}$  is a subset of  $\mathbb{R}^{k \times 2}$  containing a unique element of each orbit. A well-known example of section for the similarity group acting on  $\mathbb{R}^{k \times 2} \setminus \mathbf{1}_{k \times 2}$  is the so-called Bookstein's coordinates system (see e.g. [Dryden & Mardia, 1998] p. 27).

## 2.2 Kendall's shape space and Generalized Procrustes analysis

**Shape space** Let  $\mathbf{x} \in \mathbb{R}^{k \times 2} \setminus \mathbf{1}_{k \times 2}$  be a non-degenerated configuration. Let  $H = Id_k - \frac{1}{k} \mathbf{1}_k \mathbf{1}_k'$  be a centering matrix. The effect of translation can be eliminated by centering the configuration  $\mathbf{x}$  using the matrix  $H$  (see [Dryden & Mardia, 1998] for other centering methods), while the effect of isotropic scaling is removed by projecting the centered configuration on a unit sphere, which yields to the so-called pre-shape  $\mathbf{x}^0$  of  $\mathbf{x}$  defined as

$$\mathbf{x}^0 = \frac{H\mathbf{x}}{\|H\mathbf{x}\|_{\mathbb{R}^{k \times 2}}} \in \mathbb{R}^{k \times 2}.$$

Consider now the pre-shape sphere defined by  $\mathbb{S}_2^k := \{\mathbf{x}^0, \mathbf{x} \in \mathbb{R}^{k \times 2} \setminus \mathbf{1}_{k \times 2}\}$  and see Figure 1 for an illustration. Note that this normalization of the planar configurations amounts to choose a section for

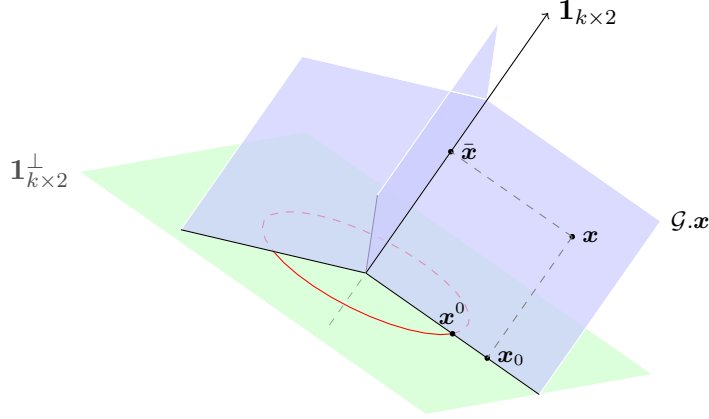


Figure 1: Three orbits of the action of the similarity group  $\mathcal{G}$  are represented in blue. The space of centered configurations  $\mathbf{1}_{k \times 2}^\perp$  is the green plane. The preshape sphere  $\mathbb{S}_2^k$  is the red circle. For a particular  $\mathbf{x} \in \mathbb{R}^{k \times 2}$ , the centered version is  $\mathbf{x}_0$  and the centered and normalized version is  $\mathbf{x}^0$ . The degenerated configuration associated to  $\mathbf{x}$  is  $\bar{\mathbf{x}}$ .

the action of the group generated by the translation and scaling in the plane. The Kendall's shape space is then defined as the quotient of  $\mathbb{S}_2^k$  by the group  $\mathbf{SO}(2)$  of rotations of the plane, namely

$$\Sigma_2^k := \mathbb{S}_2^k / \mathbf{SO}(2) = \left\{ [\mathbf{x}^0] : \mathbf{x}^0 \in \mathbb{S}_2^k \right\} \text{ with } [\mathbf{x}^0] = \{ \mathbf{x}^0 R_\alpha, \alpha \in [-\pi, \pi[ \}.$$

The space  $\Sigma_2^k$  can be endowed with a Riemannian structure and we refer to [Kendall et al., 1999] for a detail discussion on its geometric properties.

Let us briefly recall the definition of the so-called partial and full Procrustes distances on  $\mathbb{S}_2^k$ . The partial Procrustes distance is defined on the pre-shape sphere  $\mathbb{S}_2^k$  as

$$d_P^2(\mathbf{x}^0, \mathbf{y}^0) = \inf_{\alpha \in [-\pi, \pi[} \|\mathbf{x}^0 - \mathbf{y}^0 R_\alpha\|_{\mathbb{R}^{k \times 2}}^2, \quad \mathbf{x}, \mathbf{y} \in \mathbb{S}_2^k.$$

Hence, it is the (Euclidean) distance between the orbits  $[\mathbf{x}^0] = \mathbf{SO}(2) \cdot \mathbf{x}^0$  and  $[\mathbf{y}^0] = \mathbf{SO}(2) \cdot \mathbf{y}^0$  with  $\mathbf{x}^0, \mathbf{y}^0 \in \mathbb{S}_2^k$ . Let now  $\mathcal{H}$  be the group of transformations of the plane generated by scaling and rotations. The action of  $h \in \mathcal{H}$  on the centered configuration  $\mathbf{x}^0$  is defined as  $h \cdot \mathbf{x}^0 := e^a \mathbf{x}^0 R_\alpha$  where  $h = (a, \alpha) \in \mathbb{R} \times [-\pi, \pi[$ . The Full Procrustes distance is then defined as

$$d_F^2(\mathbf{x}^0, \mathbf{y}^0) = \inf_{h \in \mathcal{H}} \|\mathbf{x}^0 - h \cdot \mathbf{y}^0\|_{\mathbb{R}^{k \times 2}}^2, \quad \mathbf{x}, \mathbf{y} \in \mathbb{S}_2^k.$$

**Generalized Procrustes analysis** The full Procrustes sample mean  $\hat{\mathbf{Y}}_F$  of  $\mathbf{Y}_1, \dots, \mathbf{Y}_J$  (see e.g. [Goodall, 1991, Dryden & Mardia, 1998]) is defined by

$$\hat{\mathbf{Y}}_F = \operatorname{argmin}_{\mathbf{x}^0 \in \mathbb{S}_2^k} \sum_{j=1}^J d_F^2(\mathbf{Y}_j^0, \mathbf{x}^0). \quad (2.2)$$

The partial Procrustes mean  $\hat{\mathbf{Y}}_P$  is defined in the same way by replacing  $d_F$  by  $d_P$  in (2.2). Thence, this two Procrustes means are Fréchet mean either on  $(\mathbb{S}_2^k, d_F)$  or  $(\mathbb{S}_2^k, d_P)$  endowed with the empirical measure  $\mu_J = \sum_{j=1}^J \delta_{\mathbf{Y}_j}$ .

In practice, there are several way to compute the full Procrustes mean. In [Kent, 1992] the author used complex coordinates and expressed the full Procrustes mean as the biggest eigenvalue of a symmetrical positive definite complex matrix, see [Dryden & Mardia, 1998] result 3.2 and [Bhattacharya & Patrangenaru, 2001].



The full Procrustes mean  $\hat{\mathbf{Y}}_F$  can also be approximated by General Procrustes procedure which amounts to use the following identity

$$\hat{\mathbf{Y}}_F = \frac{1}{J} \sum_{j=1}^J \hat{h}_j \cdot \mathbf{Y}_j^0 \quad (2.3)$$

where  $\hat{h}_1, \dots, \hat{h}_J$  are the argmins of the functional  $M(h_1, \dots, h_J) = \frac{1}{J} \sum_{j=1}^J \|h_j \cdot \mathbf{Y}_j^0 - \frac{1}{J} \sum_{j'=1}^J h_{j'} \cdot \mathbf{Y}_{j'}^0\|_{\mathbb{R}^{k \times 2}}^2$  subject to the constraint  $\|\frac{1}{J} \sum_{j=1}^J h_j \cdot \mathbf{Y}_j^0\|_{\mathbb{R}^{k \times 2}}^2 = 1$ . The configurations  $\hat{h} \cdot \mathbf{Y}_j$ 's are known as Full Procrustes fits and the  $\hat{h}$  can be explicitly computed by using a singular value decomposition. In practice, one can use the iterative General Procrustes algorithm to compute  $\hat{\mathbf{Y}}$ , see [Dryden & Mardia, 1998] pages 90-91.

## 2.3 Discussion on the double asymptotic setting and comparison with GPA

**Asymptotic settings** When random planar curves (such as digits or letters for instance) are observed, a natural framework for statistical inference is an asymptotic setting in the number  $J$  of curves. This setting means that increasing the number of curves at hand should help to compute a more accurate empirical mean curve. Unfortunately, consistency results of Procrustes type procedures are reduced to the very specific case of an isotropic perturbation. In this paper, we show that increasing the number  $k$  of discretization points will ensure a consistent estimation of a mean shape in more general cases.

Consider model (1.1) where  $k > 2$  is fixed and the random perturbation  $\zeta$  is isotropic, (see Proposition 1 in [Le, 1998] for a precise definition of isotropy for random variables belonging to  $\mathbb{R}^{k \times 2}$ ). In this framework, it has been proved in [Kent & Mardia, 1997] that the functional  $M_J(\mathbf{x}^0) = \frac{1}{J} \sum_{j=1}^J d_F^2(\mathbf{Y}_j, \mathbf{x}^0)$  defined on  $\mathbb{R}^{k \times 2}$  converge uniformly in probability to the functional  $M(\mathbf{x}^0) = \mathbb{E}(d_F^2(\mathbf{Y}_1, \mathbf{x}^0))$  which admits a unique minimum at  $[\mathbf{f}^0]$ . These two facts imply that  $[\hat{\mathbf{Y}}_F]$  converges almost surely to  $[\mathbf{f}^0]$  as  $J \rightarrow +\infty$ . In [Le, 1998] the author used a Fréchet mean approach to show the consistency of  $\hat{\mathbf{Y}}_F$  with a slightly more general kind of noise. Finally, note that the estimator  $\hat{\mathbf{Y}}_P$  defined above is also studied in [Kent & Mardia, 1997, Le, 1998] with similar consistency results.

When the random perturbation  $\zeta$  in model (1.1) is non-isotropic, it has been argued in [Kent & Mardia, 1997] that the Procrustes estimator  $\hat{\mathbf{Y}}_F$  can be arbitrarily inconsistent when the signal-to-noise ratio decrease. The heuristic presented by the authors suggests that the main phenomenon that prevent the Procrustes estimator to be consistent is the fact that the functional  $M$  do not attain its minimum at  $[\mathbf{f}^0]$ . In section 4.2 of [Huckemann, 2011] the author makes this remark clear as he gives an explicit example : given a mean pattern  $\mathbf{f}^0$ , the idea is to increase the level of noise  $\zeta$  until the argmin of the functional  $M$  which was initially equals to  $[\mathbf{f}_0]$  jumps abruptly to another point. This phenomenon seems to be linked to the geometry of the sphere and properties of the Fréchet mean.

**Comparison with GPA** Hence, it is commonly the belief that Procrustes sample means can be inconsistent when considering convergence in  $\Sigma_2^k$  and the asymptotic setting  $J \rightarrow +\infty$ . Nevertheless, the above discussion suggests that a sufficient condition to ensure the consistency of Procrustes type estimators is to control the level of non-isotropic noise. That is why we introduced a pre-smoothing step that takes advantage of increasing the number  $k$  of points composing each configuration in order to ensure more general consistency results. Therefore, our approach and GPA share some similarities. They are both based on the estimation of scaling, rotation and translation parameters by a Procrustean procedure which leads to the M-estimators (1.6) which is related to (2.2). To compute a sample mean shape, this M-estimation step is then followed by a standard empirical mean in  $\mathbb{R}^{k \times 2}$  of the aligned data using these estimated deformation parameters, see equations (1.8) and (2.3).

However, one of the main differences between the approach developed in this paper and GPA is the choice of the normalization of the data. In GPA, the deformation parameters  $\hat{h}_1, \dots, \hat{h}_J$  are computed so that the full Procrustes sample mean  $\hat{\mathbf{Y}}_F$  belongs to the pre-shape sphere  $\mathbb{S}_2^k$ , see the

constraint appearing in (2.2). Therefore, the computation of  $\hat{h}_1, \dots, \hat{h}_J$  is somewhat independent of any assumption on the true parameters  $(\mathbf{a}^*, \boldsymbol{\alpha}^*, \mathbf{b}^*)$  in model (1.1). In this paper, to ensure the well-posedness of the problem (1.6), we chose to compute the estimator  $(\hat{\mathbf{a}}^\lambda, \hat{\boldsymbol{\alpha}}^\lambda, \hat{\mathbf{b}}^\lambda)$  by minimizing the matching criterion (1.5) on the constrained set  $\Theta_0$ . The choice of the constraints in  $\Theta_0$  is motivated by the hypothesis that the true deformation parameters  $(\mathbf{a}^*, \boldsymbol{\alpha}^*, \mathbf{b}^*)$  in (1.1) are centered around the identity.

### 3 Identifiability conditions

Recall that in model (1.1), the random deformations acting on the the mean pattern  $\mathbf{f}$  are parametrized by a vector  $(\mathbf{a}^*, \boldsymbol{\alpha}^*, \mathbf{b}^*) = (a_1^*, \dots, a_J^*, \alpha_1^*, \dots, \alpha_J^*, b_1^*, \dots, b_J^*)$  in  $\mathbb{R}^J \times [-\pi, \pi]^J \times \mathbb{R}^{2J}$ .

**Assumption 3.** Let  $0 < A, B$  and  $0 < \mathcal{A} < \pi$  be three real numbers. The deformation parameters  $(\mathbf{a}_j^*, \alpha_j^*, \mathbf{b}_j^*)$ , are i.i.d random variables with zero expectation and taking their values in

$$\Theta^* = \left[-\frac{A}{2}, \frac{A}{2}\right] \times \left[-\frac{\mathcal{A}}{2}, \frac{\mathcal{A}}{2}\right] \times [-B, B]^2.$$

Let  $\Theta^* = \left[-\frac{A}{2}, \frac{A}{2}\right]^J \times \left[-\frac{\mathcal{A}}{2}, \frac{\mathcal{A}}{2}\right]^J \times [-B, B]^{2J}$ . Under Assumption 3, we have  $(\mathbf{a}^*, \boldsymbol{\alpha}^*, \mathbf{b}^*) \in \Theta^*$ . Note that the compactness of  $\Theta^*$  (and thus of  $\Theta_0$ ) is an essential condition to ensure the consistency of our procedure. Indeed, the estimation of the deformation parameters  $(\mathbf{a}^*, \boldsymbol{\alpha}^*, \mathbf{b}^*)$  and the mean pattern  $\mathbf{f}$  is based on the minimization of the criterion (1.5). If there were no restriction on the amplitude of the scaling parameter, the degenerate solution  $a_j = -\infty$  for all  $j = 1, \dots, J$  is always a minimizer of (1.5). Therefore, the minimization has to be performed under additional compact constraints.

#### 3.1 The deterministic criterion $D$

Let  $(\mathbf{a}, \boldsymbol{\alpha}, \mathbf{b}) \in \mathbb{R}^J \times [-\pi, \pi]^J \times \mathbb{R}^{2J}$  and consider the following criterion,

$$D(\mathbf{a}, \boldsymbol{\alpha}, \mathbf{b}) = \frac{1}{Jk} \sum_{j=1}^J \left\| (g_j^{-1} \cdot g_j^*) \cdot \mathbf{f} - \frac{1}{J} \sum_{j'=1}^J (g_{j'}^{-1} \cdot g_{j'}^*) \cdot \mathbf{f} \right\|_{\mathbb{R}^{k \times 2}}^2. \quad (3.1)$$

where  $g_j = (a_j, \alpha_j, b_j)$  and  $g_j^* = (a_j^*, \alpha_j^*, b_j^*)$  for all  $j = 1, \dots, J$ . The criterion  $D$  is a version without noise of the criterion  $M^\lambda$  defined at (1.5). The estimation procedure described in Section 1.1 is based on the convergence of the argmins of  $M^\lambda$  toward the argmin of  $D$  when  $k$  goes to infinity. As a consequence, choosing identifiability conditions amounts to fix a subset  $\Theta_0$  of  $\mathbb{R}^J \times [-\pi, \pi]^J \times \mathbb{R}^{2J}$  on which  $D$  has a unique argmin. In the rest of this section, we determine the zeros of  $D$ , and then we fix a convenient constraint set  $\Theta_0$  that contains a unique point at which  $D$  vanishes.

The criterion  $D$  clearly vanishes at  $(\mathbf{a}^*, \boldsymbol{\alpha}^*, \mathbf{b}^*) \in \mathbb{R}^J \times [-\pi, \pi]^J \times \mathbb{R}^{2J}$ . This minimum is not unique since easy algebra implies that

$$D(\mathbf{a}, \boldsymbol{\alpha}, \mathbf{b}) = 0 \quad \Longleftrightarrow \quad (g_j^{-1} \cdot g_j^*) \cdot \mathbf{f} = (g_{j'}^{-1} \cdot g_{j'}^*) \cdot \mathbf{f}, \quad \text{for all } j, j' = 1, \dots, J.$$

Suppose now that  $\mathbf{f} \notin \mathbf{1}_{2 \times k}$  is a non-degenerated planar configuration. In Section 2.1, we have seen that the action of  $\mathcal{G}$  on  $\mathbf{f}$  is free, that is, the stabilizer  $I(\mathbf{f})$  is reduced to the identity. Thus, we obtain,

$$\begin{aligned} D(\mathbf{a}, \boldsymbol{\alpha}, \mathbf{b}) = 0 & \quad \Longleftrightarrow \quad g_j^{-1} \cdot g_j^* = g_{j'}^{-1} \cdot g_{j'}^*, & \text{for all } j, j' = 1, \dots, J \\ & \Longleftrightarrow \quad \begin{cases} a_j^* - a_j = a_{j'}^* - a_{j'}, \\ \alpha_j^* - \alpha_j = \alpha_{j'}^* - \alpha_{j'}, \\ e^{-a_j} (b_j^* - b_j) R_{-\alpha_j} = e^{-a_{j'}} (b_{j'}^* - b_{j'}) R_{-\alpha_{j'}}, \end{cases} \\ & \text{for all } j, j' = 1, \dots, J \end{aligned}$$

We have proved the following result,

**Lemma 3.1.** Let  $\mathbf{f} \in \mathbb{R}^{k \times 2}$  be a non-degenerated configuration of  $k$ -ads in the plane, i.e.  $\mathbf{f} \notin \mathbf{1}_{2 \times k}$ . Then,  $D(\mathbf{a}, \boldsymbol{\alpha}, \mathbf{b}) = 0$  if and only if  $(\mathbf{a}, \boldsymbol{\alpha}, \mathbf{b})$  belongs to the set

$$(\mathbf{a}^*, \boldsymbol{\alpha}^*, \mathbf{b}^*) * \mathcal{G} = \left\{ (\mathbf{a}^*, \boldsymbol{\alpha}^*, \mathbf{b}^*) * (a_0, \alpha_0, b_0), (a_0, \alpha_0, b_0) \in \mathbb{R} \times [-\pi, \pi[\times \mathbb{R}^2] \right\},$$

where  $(\mathbf{a}^*, \boldsymbol{\alpha}^*, \mathbf{b}^*) * (a_0, \alpha_0, b_0) = (a_1^* + a_0, \dots, a_J^* + a_0, \alpha_1^* + \alpha_0, \dots, \alpha_J^* + \alpha_0, e^{a_1^*} b_0 R_{\alpha_1^*} + b_1^*, \dots, e^{a_J^*} b_0 R_{\alpha_J^*} + b_J^*) \in \mathbb{R}^J \times [-\pi, \pi[ \times \mathbb{R}^{2J}$ .

**Remark 1.** Lemma 3.1 is simpler than it appears. By reordering the entries of the vector  $(\mathbf{a}^*, \boldsymbol{\alpha}^*, \mathbf{b}^*)$  there is an obvious correspondence between  $(\mathbf{a}^*, \boldsymbol{\alpha}^*, \mathbf{b}^*) \in \boldsymbol{\Theta}^*$  and  $(g_1^*, \dots, g_J^*) \in \mathcal{G}^J$  via the parametrization of the similarity group defined in Section 2.1. Hence, Lemma 3.1 tells us that the criterion  $D$  vanishes for all the vectors  $(\mathbf{a}, \boldsymbol{\alpha}, \mathbf{b}) \in \mathbb{R}^J \times [-\pi, \pi[ \times \mathbb{R}^{2J}$  corresponding to the subset of the group  $\mathcal{G}^J$  given by

$$(g_1^*, \dots, g_J^*) * \mathcal{G} = \{(g_1^* \cdot g_0, \dots, g_J^* \cdot g_0), g_0 \in \mathcal{G}\} \subset \mathcal{G}^J.$$

The “\*” notation is nothing else than the right composition by a same  $g_0 \in \mathcal{G}$  of all the entries of a  $(g_1, \dots, g_J) \in \mathcal{G}^J$ . Hence the subset  $(g_1^*, \dots, g_J^*) * \mathcal{G}$  can be interpreted as the orbit of  $(g_1^*, \dots, g_J^*) \in \mathcal{G}^J$  under the (right) action of  $\mathcal{G}$ . Indeed,  $\mathcal{G}$  acts naturally by (right) composition on all the coordinates of an element of  $\mathcal{G}^J$ .

### 3.2 The constraint set $\boldsymbol{\Theta}_0$

By Lemma 3.1, the set  $\boldsymbol{\Theta}_0$  must intersect at a unique point, say  $(\mathbf{a}_{\boldsymbol{\Theta}_0}^*, \boldsymbol{\alpha}_{\boldsymbol{\Theta}_0}^*, \mathbf{b}_{\boldsymbol{\Theta}_0}^*)$ , each set  $(\mathbf{a}^*, \boldsymbol{\alpha}^*, \mathbf{b}^*) * \mathcal{G}$ . It is convenient to choose  $\boldsymbol{\Theta}_0$  to be of the form  $\boldsymbol{\Theta}_0 = \mathbb{R}^J \times [-\pi, \pi[ \times \mathbb{R}^{2J} \cap \mathcal{L}_0$  where  $\mathcal{L}_0$  is a linear space of  $\mathbb{R}^{4J}$ . The linear space  $\mathcal{L}_0$  must be chosen so that for any  $(\mathbf{a}^*, \boldsymbol{\alpha}^*, \mathbf{b}^*)$  in  $\boldsymbol{\Theta}^*$ , there exists a unique point  $(\mathbf{a}_{\boldsymbol{\Theta}_0}^*, \boldsymbol{\alpha}_{\boldsymbol{\Theta}_0}^*, \mathbf{b}_{\boldsymbol{\Theta}_0}^*)$  in  $\boldsymbol{\Theta}_0$  that can be written as  $(\mathbf{a}_{\boldsymbol{\Theta}_0}^*, \boldsymbol{\alpha}_{\boldsymbol{\Theta}_0}^*, \mathbf{b}_{\boldsymbol{\Theta}_0}^*) = (\mathbf{a}^*, \boldsymbol{\alpha}^*, \mathbf{b}^*) * (a_0, \alpha_0, b_0)$  for some  $(a_0, \alpha_0, b_0) \in \mathbb{R} \times [-\pi, \pi[ \times \mathbb{R}^2$ .

**Remark 2.** As we have seen in Remark 1, the set  $(\mathbf{a}^*, \boldsymbol{\alpha}^*, \mathbf{b}^*) * \mathcal{G}$  can be interpreted as an orbit of the action of  $\mathcal{G}$  on  $\mathcal{G}^J$ . In this terminology, the set  $\boldsymbol{\Theta}_0$  can be viewed as a section of the orbits. Indeed, the section is the set of representatives  $(\mathbf{a}_{\boldsymbol{\Theta}_0}^*, \boldsymbol{\alpha}_{\boldsymbol{\Theta}_0}^*, \mathbf{b}_{\boldsymbol{\Theta}_0}^*)$  of each orbit. See Figure 2 for an illustration.

Let us consider a choice of  $\boldsymbol{\Theta}_0$  motivated by the fact that, under Assumption 3, the random deformation parameters have zero expectation. In this setting, it is natural to impose that the estimated deformation parameters sum up to zero by choosing  $\mathcal{L}_0 = \mathbf{1}_{4J}^\perp$ , which is the orthogonal of the linear space  $\mathbf{1}_{4J} = \mathbb{1}_{4J} \cdot \mathbb{R} \subset \mathbb{R}^{4J}$ . Such a choice leads to the set  $\boldsymbol{\Theta}_0$  defined equation (1.7) i.e.

$$\boldsymbol{\Theta}_0 = \{(\mathbf{a}, \boldsymbol{\alpha}, \mathbf{b}) \in \boldsymbol{\Theta}^J, (a_1 + \dots + a_J, \alpha_1 + \dots + \alpha_J, b_1 + \dots + b_J) = 0\}.$$

Now, let us show that for any  $(\mathbf{a}^*, \boldsymbol{\alpha}^*, \mathbf{b}^*) \in \boldsymbol{\Theta}^*$  there exists a unique  $(\mathbf{a}_{\boldsymbol{\Theta}_0}^*, \boldsymbol{\alpha}_{\boldsymbol{\Theta}_0}^*, \mathbf{b}_{\boldsymbol{\Theta}_0}^*) = (\mathbf{a}^*, \boldsymbol{\alpha}^*, \mathbf{b}^*) * (a_0, \alpha_0, b_0) \in \boldsymbol{\Theta}_0$  for some  $(a_0, \alpha_0, b_0) \in \mathbb{R} \times [-\pi, \pi[ \times \mathbb{R}^2$ . This amounts to solve the following equations

$$\begin{cases} a_1^* + a_0 & + & \dots & + & a_J^* + a_0 & = & 0, \\ \alpha_1^* + \alpha_0 & + & \dots & + & \alpha_J^* + \alpha_0 & = & 0, \\ e^{a_1^*} b_0 R_{\alpha_1^*} + b_1^* & + & \dots & + & e^{a_J^*} b_0 R_{\alpha_J^*} + b_J^* & = & 0. \end{cases} \quad (3.2)$$

After some computations, we obtain that equations (3.2) are satisfied if and only if

$$(a_0, \alpha_0, b_0) = (-\bar{\mathbf{a}}^*, -\bar{\boldsymbol{\alpha}}^*, -\bar{\mathbf{b}}^* (\overline{e^{\mathbf{a}^*} R_{\boldsymbol{\alpha}^*}})^{-1}), \quad (3.3)$$

where  $\bar{\mathbf{a}}^* = \frac{1}{J} \sum_{j=1}^J a_j^* \in \mathbb{R}$ ,  $\bar{\boldsymbol{\alpha}}^* = \frac{1}{J} \sum_{j=1}^J \alpha_j^* \in \mathbb{R}$ ,  $\bar{\mathbf{b}}^* = \frac{1}{J} \sum_{j=1}^J b_j^* \in \mathbb{R}^2$  and  $\overline{e^{\mathbf{a}^*} R_{\boldsymbol{\alpha}^*}} = \frac{1}{J} \sum_{j=1}^J e^{a_j^*} R_{\alpha_j^*}$  is a  $2 \times 2$  invertible matrix. Therefore,  $(\mathbf{a}_{\boldsymbol{\Theta}_0}^*, \boldsymbol{\alpha}_{\boldsymbol{\Theta}_0}^*, \mathbf{b}_{\boldsymbol{\Theta}_0}^*)$  is uniquely given by

$$([\mathbf{a}_{\boldsymbol{\Theta}_0}^*]_j, [\boldsymbol{\alpha}_{\boldsymbol{\Theta}_0}^*]_j, [\mathbf{b}_{\boldsymbol{\Theta}_0}^*]_j) = (a_j^* - \bar{\mathbf{a}}^*, \alpha_j^* - \bar{\boldsymbol{\alpha}}^*, b_j^* - e^{a_j^*} (\bar{\mathbf{b}}^* (\overline{e^{\mathbf{a}^*} R_{\boldsymbol{\alpha}^*}})^{-1}) R_{\alpha_j^*}),$$

for  $j = 1, \dots, J$ .

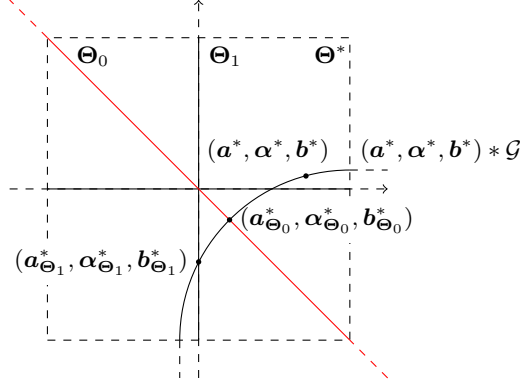


Figure 2: Choice of identifiability conditions when  $J = 2$ .

**Remark 3.** Another possible approach is to fix, say the first observation as a reference, meaning that the criterion  $D$  could be optimized on the following subspace of  $\mathbb{R}^J \times [-\pi, \pi]^J \times \mathbb{R}^{2J}$

$$\Theta_1 = \{(\mathbf{a}, \boldsymbol{\alpha}, \mathbf{b}) \in [-A, A]^J \times [-\mathcal{A}, \mathcal{A}]^{2J} \times \mathbb{R}^2, (a_1, \alpha_1, b_1) = 0\}.$$

With such a choice, for any  $(\mathbf{a}^*, \boldsymbol{\alpha}^*, \mathbf{b}^*) \in \Theta^J$ , the  $j$ -th coordinate of  $(\mathbf{a}_{\Theta_1}^*, \boldsymbol{\alpha}_{\Theta_1}^*, \mathbf{b}_{\Theta_1}^*) = (\mathbf{a}^*, \boldsymbol{\alpha}^*, \mathbf{b}^*) * (a_0, \alpha_0, b_0)$  is given by

$$([\mathbf{a}_{\Theta_1}^*]_j, [\boldsymbol{\alpha}_{\Theta_1}^*]_j, [\mathbf{b}_{\Theta_1}^*]_j) = (a_j^* - a_1^*, \alpha_j^* - \alpha_1^*, b_j^* - e^{a_j^* - a_1^*} b_1^* R_{\alpha_j^* - \alpha_1^*}),$$

where  $(a_0, \alpha_0, b_0) = (a_1^*, \alpha_1^*, b_1^*)^{-1} = (-a_1^*, -\alpha_1^*, -e^{-a_1^*} b_1^* R_{-\alpha_1^*})$ . A graphical illustration of the choice of identifiability conditions for  $J = 2$  is given in Figure 2.

## 4 The estimating procedure

### 4.1 A dimension reduction step

We use Fourier filtering to project the data into a low-dimensional space as follows. Assume for convenience that  $k$  is odd. For  $\mathbf{x} = (\mathbf{x}'_1, \dots, \mathbf{x}'_k)' \in \mathbb{R}^{k \times 2}$  and  $m = -\frac{k-1}{2}, \dots, \frac{k-1}{2}$ , let

$$c_m(\mathbf{x}) = \sum_{\ell=1}^k \mathbf{x}_\ell e^{-i2\pi m \frac{\ell}{k}} = \left( \sum_{\ell=1}^k x_\ell^{(1)} e^{-i2\pi m \frac{\ell}{k}}, \sum_{\ell=1}^k x_\ell^{(2)} e^{-i2\pi m \frac{\ell}{k}} \right) \in \mathbb{C}^2 \text{ with } \mathbf{x}_\ell = (x_\ell^{(1)}, x_\ell^{(2)}),$$

be the  $m$ -th (discrete) Fourier coefficient of  $\mathbf{x}$ . Let  $\lambda \in \{1, \dots, \frac{k-1}{2}\}$  be a smoothing parameter, and define for each  $\mathbf{Y}_j$  the smoothed shapes

$$\hat{\mathbf{f}}_j^\lambda = \left( \frac{1}{k} \sum_{0 \leq |m| \leq \lambda} c_m(\mathbf{Y}_j) e^{i2\pi m \frac{\ell}{k}} \right)_{\ell=1}^k = \mathbf{A}^\lambda \mathbf{Y}_j \in \mathbb{R}^{k \times 2}.$$

In Section 2.1, we have shown that the similarity group is not free on the subset  $\mathbf{1}_{k \times 2}$  of degenerated configurations composed of  $k$  identical landmarks, see Section 2.1. That is why we are going to treat separately the subspace  $\mathbf{1}_{k \times 2}$  and  $\mathbf{1}_{k \times 2}^\perp$  by considering the matrices

$$\bar{A} = \frac{1}{k} \mathbf{1}_k \mathbf{1}_k' \quad \text{and} \quad A_0^\lambda = \left( \frac{1}{k} \sum_{0 < |m| \leq \lambda} e^{i2\pi m \frac{\ell - \ell'}{k}} \right)_{\ell, \ell'=1}^k. \quad (4.1)$$

Remark that  $\bar{A}$  is a projection matrix on the one dimensional sub-space  $\bar{\mathcal{V}} := \mathbf{1}_k \cdot \mathbb{R} = \{c \mathbf{1}_k : c \in \mathbb{R}\}$  of  $\mathbb{R}^k$ . The matrix  $A_0^\lambda$  is a projection matrix in a (trigonometric) sub-space  $\mathcal{V}_0^\lambda$  of dimension  $2\lambda$ . Note

that it is included in the linear space  $\bar{\mathcal{V}}^\perp = \{x \in \mathbb{R}^k : x' \mathbf{1}_k = 0\}$ . Hence,  $\mathcal{V}_0^\lambda \times \mathcal{V}_0^\lambda$  is a linear subspace of  $\mathbf{1}_{k \times 2}^\perp$  which is the space of the centered configurations. We have,

$$\mathbf{A}^\lambda = A_0^\lambda + \bar{A} \quad \text{and} \quad \mathcal{V}^\lambda = \mathcal{V}_0^\lambda \oplus \bar{\mathcal{V}}. \quad (4.2)$$

Thus, we can write the smoothed shape  $\hat{\mathbf{f}}_j^\lambda$  as

$$\hat{\mathbf{f}}_j^\lambda = \mathbf{A}^\lambda \mathbf{Y}_j = A_0^\lambda \mathbf{Y}_j + \bar{A} \mathbf{Y}_j \in (\bar{\mathcal{V}} \times \bar{\mathcal{V}}) \oplus (\mathcal{V}_0^\lambda \times \mathcal{V}_0^\lambda)$$

where  $\bar{\mathcal{V}} \times \bar{\mathcal{V}} = \mathbf{1}_{k \times 2}$  and  $\mathcal{V}_0^\lambda \times \mathcal{V}_0^\lambda \subset \mathbf{1}_{k \times 2}^\perp$ . In other words,  $A_0^\lambda \mathbf{Y}_j$  is the smoothed centered configuration associated to  $\mathbf{Y}_j$  and  $\bar{A} \mathbf{Y}_j$  is the degenerated configuration given by the Euclidean mean of the  $k$  landmarks composing  $\mathbf{Y}_j$ . Finally, remark that the low pass filter and the action of similarity group commute, that is, we have for all  $g \in \mathcal{G}$  and  $\mathbf{f} \in \mathbb{R}^{k \times 2}$

$$g \cdot (\mathbf{A}^\lambda \mathbf{f}) = e^a \mathbf{A}^\lambda \mathbf{f} R_\alpha + \mathbf{1}_k \otimes b = \mathbf{A}^\lambda (e^a \mathbf{f} R_\alpha + \mathbf{1}_k \otimes b) = \mathbf{A}^\lambda (g \cdot \mathbf{f}).$$

## 4.2 Estimation of the deformation parameters

Recall that the estimator  $(\hat{\mathbf{a}}^\lambda, \hat{\boldsymbol{\alpha}}^\lambda, \hat{\mathbf{b}}^\lambda)$  of  $(\mathbf{a}^*, \boldsymbol{\alpha}^*, \mathbf{b}^*)$  is defined by the optimization problem (1.6). Nevertheless, as it is suggested by the discussion of Sections 2.1 and 4.1, one can carry out the estimation process in two steps. First, we estimate the rotation and scaling parameters on the space  $\mathcal{V}_0^\lambda \times \mathcal{V}_0^\lambda \subset \mathbf{1}_{k \times 2}^\perp$  of the centered configurations. We then use these estimators to estimate the translation parameters which act on  $\bar{\mathcal{V}} \times \bar{\mathcal{V}} = \mathbf{1}_{k \times 2}$ . Note that this procedure is equivalent to the optimization problem (1.6) as shown by Lemma 4.1 below.

**Estimation of rotations and scaling.** Define

$$M_0^\lambda(\mathbf{a}, \boldsymbol{\alpha}) = \frac{1}{Jk} \sum_{j=1}^J \left\| e^{-a_j} A_0^\lambda \mathbf{Y}_j R_{-\alpha_j} - \frac{1}{J} \sum_{j'=1}^J e^{-a_{j'}} A_0^\lambda \mathbf{Y}_{j'} R_{-\alpha_{j'}} \right\|_{\mathbb{R}^{k \times 2}}^2.$$

Let  $\mathbf{1}_J^\perp = \{(a_1, \dots, a_J) \in \mathbb{R}^J, a_1 + \dots, a_J = 0\}$  and consider the space  $\boldsymbol{\Theta}_0^{\mathbf{a}, \boldsymbol{\alpha}} = ([-A, A]^J \cap \mathbf{1}_J^\perp) \times ([-\mathcal{A}, \mathcal{A}]^J \cap \mathbf{1}_J^\perp)$ . Then, estimators of the rotation and scaling parameters are given by

$$(\hat{\mathbf{a}}^\lambda, \hat{\boldsymbol{\alpha}}^\lambda) \in \underset{(\mathbf{a}, \boldsymbol{\alpha}) \in \boldsymbol{\Theta}_0^{\mathbf{a}, \boldsymbol{\alpha}}}{\operatorname{argmin}} M_0^\lambda(\mathbf{a}, \boldsymbol{\alpha}). \quad (4.3)$$

**Estimation of translations.** Now that we have computed estimators of the rotation and scaling parameters, let us define the criterion,

$$\bar{M}(\mathbf{a}, \boldsymbol{\alpha}, \mathbf{b}) = \frac{1}{Jk} \sum_{j=1}^J \left\| e^{-a_j} \bar{A}(\mathbf{Y}_j - \mathbf{1}_k \otimes b_j) R_{-\alpha_j} - \frac{1}{J} \sum_{j'=1}^J e^{-a_{j'}} \bar{A}(\mathbf{Y}_{j'} - \mathbf{1}_k \otimes b_{j'}) R_{-\alpha_{j'}} \right\|_{\mathbb{R}^{k \times 2}}^2$$

and the space  $\boldsymbol{\Theta}_0^{\mathbf{b}} = \{(b_1^{(1)}, \dots, b_J^{(1)}, b_1^{(2)}, \dots, b_J^{(2)}) \in \mathbb{R}^{2J}, b_1^{(1)} + \dots + b_J^{(1)} = b_1^{(2)} + \dots + b_J^{(2)} = 0\}$ . The estimator of the translation parameters is then given by,

$$\hat{\mathbf{b}}^\lambda = \underset{\mathbf{b} \in \boldsymbol{\Theta}_0^{\mathbf{b}}}{\operatorname{argmin}} \bar{M}(\hat{\mathbf{a}}^\lambda, \hat{\boldsymbol{\alpha}}^\lambda, \mathbf{b}). \quad (4.4)$$

We emphasize that the estimators of the translation parameters depend on the estimated rotation and scaling parameters. It is shown in the proof of Lemma 4.1 below that we have an explicit expression of  $\hat{\mathbf{b}}^\lambda$  given by

$$\hat{\mathbf{b}}^\lambda = (\bar{\mathbf{Y}}_1 - e^{a_1} \bar{\mathbf{Y}} (\overline{e^a R_\alpha})^{-1} R_{\alpha_1}, \dots, \bar{\mathbf{Y}}_J - e^{a_J} \bar{\mathbf{Y}} (\overline{e^a R_\alpha})^{-1} R_{\alpha_J}) \quad (4.5)$$

where  $\overline{e^{\mathbf{a}} R \boldsymbol{\alpha}} = \frac{1}{J} \sum_{j=1}^J e^{a_j} R_{\alpha_j} \in \mathbb{R}^{2 \times 2}$ ,  $\bar{\mathbf{Y}}_j = \frac{1}{k} \sum_{\ell=1}^k [\mathbf{Y}_j]_{\ell} \in \mathbb{R}^2$ , that is  $\bar{A} \mathbf{Y}_j = \mathbf{1}_k \otimes \bar{\mathbf{Y}}_j \in \mathbb{R}^{k \times 2}$  is a degenerated configuration, and  $\bar{\mathbf{Y}} = \frac{1}{J} \sum_{j=1}^J \bar{\mathbf{Y}}_j \in \mathbb{R}^2$ .

This two steps procedure is equivalent to the optimization problem (1.6) as we have the following decompositions  $M^\lambda(\mathbf{a}, \boldsymbol{\alpha}, \mathbf{b}) = M_0^\lambda(\mathbf{a}, \boldsymbol{\alpha}) + \bar{M}(\mathbf{a}, \boldsymbol{\alpha}, \mathbf{b})$  and  $\boldsymbol{\Theta}_0 = \boldsymbol{\Theta}_0^{\mathbf{a}, \boldsymbol{\alpha}} \times \boldsymbol{\Theta}_0^{\mathbf{b}}$ , implying following result (see the Appendix for a detailed proof)

$$\textbf{Lemma 4.1. } (\hat{\mathbf{a}}^\lambda, \hat{\boldsymbol{\alpha}}^\lambda, \hat{\mathbf{b}}^\lambda) \in \underset{(\mathbf{a}, \boldsymbol{\alpha}, \mathbf{b}) \in \boldsymbol{\Theta}_0}{\operatorname{argmin}} M^\lambda(\mathbf{a}, \boldsymbol{\alpha}, \mathbf{b}) \iff \begin{cases} (\hat{\mathbf{a}}^\lambda, \hat{\boldsymbol{\alpha}}^\lambda) \in \underset{(\mathbf{a}, \boldsymbol{\alpha}) \in \boldsymbol{\Theta}_0^{\mathbf{a}, \boldsymbol{\alpha}}}{\operatorname{argmin}} M_0^\lambda(\mathbf{a}, \boldsymbol{\alpha}) \\ \hat{\mathbf{b}}^\lambda = \underset{\mathbf{b} \in \boldsymbol{\Theta}_0^{\mathbf{b}}}{\operatorname{argmin}} \bar{M}(\hat{\mathbf{a}}^\lambda, \hat{\boldsymbol{\alpha}}^\lambda, \mathbf{b}). \end{cases}$$

## 5 Consistency results

In what follows,  $C, C_0, C_1$  denote positive constants whose value may change from line to line. The notation  $C(\cdot)$  specifies the dependency of  $C$  on some quantities.

### 5.1 Consistent estimation of the deformation parameters

**Rotation and scaling.** Recall that the rotation and scaling parameters are estimated separately on the smoothed and centered observations. We have the following result,

**Theorem 5.1.** *Consider model (1.1) and suppose that Assumptions 1 and 2 hold and that Assumption 3 is verified with  $\max\{A, \mathcal{A}\} < 0.1$ . Consider  $(\hat{\mathbf{a}}^\lambda, \hat{\boldsymbol{\alpha}}^\lambda)$  the estimators defined in equation (4.3). If  $\lambda = k^{\frac{1}{2s+1}}$  then for all  $x > 0$  we have*

$$\mathbb{P} \left( \frac{1}{J} \|(\hat{\mathbf{a}}^\lambda, \hat{\boldsymbol{\alpha}}^\lambda) - (\mathbf{a}^*, \boldsymbol{\alpha}^*)\|_{\mathbb{R}^{2J}}^2 \geq C_1(L, s, A, \mathcal{A}, \mathbf{f}_0) A_1(k, J, x) + C_2(A, \mathcal{A}) A_2(J, x) \right) \leq 4e^{-x}, \quad (5.1)$$

where  $C_1(L, s, A, \mathcal{A}, \mathbf{f}_0), C_2(A, \mathcal{A}) > 0$  are positive constants independent of  $k$  and  $J$ ,  $A_1(k, J, x) = F(k^{-\frac{2s}{2s+1}}) + F(V_1(k, J, x))$  with  $V_1(k, J, x) = 3\gamma_{\max}(k) k^{-\frac{2s}{2s+1}} (1 + \sqrt{2\frac{x}{Jk}} + 2\frac{x}{Jk})$  and  $A_2(J, x) = \left(\frac{x}{3J} + \sqrt{\frac{2x}{J}}\right)^2$ , where  $F(u) = u + \sqrt{u}$ , for all  $u \geq 0$ .

Remark that a direct consequence of Theorem 5.1 is the consistency of  $(\hat{\mathbf{a}}^\lambda, \hat{\boldsymbol{\alpha}}^\lambda)$  to  $(\mathbf{a}^*, \boldsymbol{\alpha}^*)$  when  $\min\{k, J\} \rightarrow \infty$ . Indeed, we have  $\lim_{u \rightarrow 0} F(u) = 0$  and under Assumption 2 and for any fixed  $J \geq 1$  and  $x > 0$ , the term  $V_1(k, J, x)$  tends to zero as  $k$  goes to infinity. Hence for any  $x > 0$  and  $J \geq 1$ , we have

$$\lim_{k \rightarrow \infty} A_1(k, J, x) = 0.$$

Under the same hypothesis as in Theorem 5.1 but without the bounds on  $A$  and  $\mathcal{A}$ , Proposition A.1 then ensures the convergence of  $(\hat{\mathbf{a}}^\lambda, \hat{\boldsymbol{\alpha}}^\lambda)$  to  $(\mathbf{a}_{\boldsymbol{\Theta}_0}^*, \boldsymbol{\alpha}_{\boldsymbol{\Theta}_0}^*)$  as  $J$  remains fixed and  $k \rightarrow \infty$ . Now, for all  $x > 0$ , we have

$$\lim_{J \rightarrow \infty} A_2(J, x) = 0.$$

Thus a double asymptotic  $\min\{k, J\} \rightarrow \infty$  ensures that  $\frac{1}{J} \|(\hat{\mathbf{a}}^\lambda, \hat{\boldsymbol{\alpha}}^\lambda) - (\mathbf{a}^*, \boldsymbol{\alpha}^*)\|_{\mathbb{R}^{2J}}^2$  tends to 0 in probability.

**Translation parameters.** We have the following result,

**Theorem 5.2.** *Consider the hypothesis and notations of Theorem 5.1 and the estimator  $\hat{\mathbf{b}}^\lambda$  given by formula (4.4). Then we have for all  $x > 0$ ,*

$$\mathbb{P} \left( \frac{1}{J} \|\hat{\mathbf{b}}^\lambda - \mathbf{b}^*\|_{\mathbb{R}^{2J}}^2 \geq C_3(L, s, A, \mathcal{A}, B, \mathbf{f}) A_3(k, J, x) + C_4(A, \mathcal{A}, B) A_2(J, x) \right) \leq 9e^{-x}, \quad (5.2)$$

where  $C_3(L, s, A, \mathcal{A}, B, \mathbf{f}), C_4(A, \mathcal{A}, B) > 0$  are positive constants independent of  $k$  and  $J$ ,  $A_3(k, J, x) = F(k^{-\frac{2s}{2s+1}}) + F(V_1(k, J, x)) + V_2(k, J, x)$  with  $V_2(k, J, x) = 4\frac{\gamma_{\max}(k)}{k}(1 + \sqrt{2\frac{x}{J}} + \frac{x}{J})$ .

Similar comments to those made after Theorem 5.1 are still valid here. For any  $J \in \mathbb{N}$  and  $x > 0$ , we have

$$\lim_{k \rightarrow \infty} A_3(k, J, x) = 0,$$

since  $V_2(k, J, x)$  tends to 0 as  $k$  goes to infinity by Assumption 2. Under the same hypothesis as in Theorem 5.2 but without the bounds on  $A$  and  $\mathcal{A}$ , Proposition A.3 ensures the convergence in probability of  $\mathbf{b}^\lambda$  to  $\mathbf{b}_{\Theta_0}^*$  with only an asymptotic in  $k$ . In the double asymptotic setting  $\min\{k, J\} \rightarrow \infty$ , Theorem 5.2 ensure the consistency of  $\hat{\mathbf{b}}^\lambda$  to the true value  $\mathbf{b}^*$  of the translation parameters.

## 5.2 Consistent estimation of the mean shape

**Theorem 5.3.** *Consider model (1.1) and suppose that Assumptions 1 and 2 hold and that Assumption 3 is verified with  $\max\{A, \mathcal{A}\} < 0.1$ . Consider the estimator  $\hat{\mathbf{f}}^\lambda$  defined in (1.8) and let  $\lambda = k^{\frac{1}{2s+1}}$ . Then, we have for all  $x > 0$ ,*

$$\mathbb{P}\left(\frac{1}{k}\|\hat{\mathbf{f}}^\lambda - \mathbf{f}\|_{\mathbb{R}^{k \times 2}}^2 \geq C(L, s, A, \mathcal{A}, B, \mathbf{f})\left(A_1(k, J, x) + A_3(k, J, x) + A_2(J, x)\right)\right) \leq 14e^{-x},$$

where  $C(L, s, A, \mathcal{A}, B, \mathbf{f}) > 0$  is a constant independent of  $k$  and  $J$ ,  $A_1(k, J, x) = F(k^{-\frac{2s}{2s+1}}) + F(V_1(k, J, x))$  with  $V_1(k, J, x) = 3\gamma_{\max}(k)k^{-\frac{2s}{2s+1}}(1 + \sqrt{2\frac{x}{Jk}} + 2\frac{x}{Jk})$ ,  $A_3(k, J, x) = V_2(k, J, x) + F(k^{-\frac{2s}{2s+1}}) + F(V_1(k, J, x))$  with  $V_2(k, J, x) = 4\frac{\gamma_{\max}(k)}{k}(1 + \sqrt{2\frac{x}{J}} + \frac{x}{J})$  and  $A_2(J, x) = \left(\sqrt{2\frac{x}{J}} + \frac{x}{3J}\right)^2$ .

The terms  $A_1(k, J, x)$ ,  $A_3(k, J, x)$  and  $A_2(J, x)$  that appear in the statement of Theorem 5.3 are the same to those appearing in Theorems 5.1 and 5.2. As a consequence, we have  $\frac{1}{k}\|\hat{\mathbf{f}}^\lambda - \mathbf{f}\|_{\mathbb{R}^{k \times 2}}^2 \rightarrow 0$  in probability when  $\min\{k, J\} \rightarrow \infty$  and Theorem 5.3 gives rates of convergences of  $\hat{\mathbf{f}}^\lambda$  to  $\mathbf{f}$  thanks to a concentration inequality.

Theorem 5.3 is similar to Theorems 5.1 and 5.2 and there is no need to assume an extra bound on  $A$  and  $\mathcal{A}$  to ensure the convergence in probability of  $\hat{\mathbf{f}}^\lambda$  when  $J$  is fixed and  $k \rightarrow \infty$ , see the discussion following Theorems 5.1 and 5.2. Let

$$\mathbf{f}_{\Theta_0} = e^{\bar{\mathbf{a}}^*}(\mathbf{f} + \mathbb{1}_k \otimes \bar{\mathbf{b}}^*(\overline{e^{\mathbf{a}^*} R_{\alpha^*}})^{-1})R_{\bar{\alpha}^*} \quad (5.3)$$

where  $\bar{\mathbf{a}}^* = \frac{1}{J} \sum_{j=1}^J \mathbf{a}_j^* \in \mathbb{R}$ ,  $\bar{\alpha}^* = \frac{1}{J} \sum_{j=1}^J \alpha_j^* \in \mathbb{R}$ ,  $\bar{\mathbf{b}}^* = \frac{1}{J} \sum_{j=1}^J \mathbf{b}_j^* \in \mathbb{R}^2$  and  $\overline{e^{\mathbf{a}^*} R_{\alpha^*}} = \frac{1}{J} \sum_{j=1}^J e^{\mathbf{a}_j^*} R_{\alpha_j^*}$  is an invertible  $2 \times 2$  matrix, see also formula (3.3). A slight modification of the proof of Theorem 5.3 gives the following inequality,

$$\mathbb{P}\left(\frac{1}{k}\|\hat{\mathbf{f}}^\lambda - \mathbf{f}_{\Theta_0}\|^2 \geq C(L, s, A, \mathcal{A}, B, \mathbf{f})\left(A_1(k, J, x) + A_3(k, J, x)\right)\right) \leq 14e^{-x}.$$

This yields to statements (1.9) in Theorem 1.1.

## 6 Numerical experiments

### 6.1 Description of the data

We make here some numerical simulations to show the effect of the dimension  $k$  and the number  $J$  of observations on the estimation of the deformation parameters and the mean pattern with data generated by model (1.1). Different types of noise are considered. For all  $t \in [0, 1]$ , let

$$f(t) = (10 \sin^2(\pi t) + \cos(10\pi) + 20, 2 \sin(6\pi t) - 11 \sin^2(\pi t) + 12 \exp(-25(t - 0.4)^2) + 1).$$

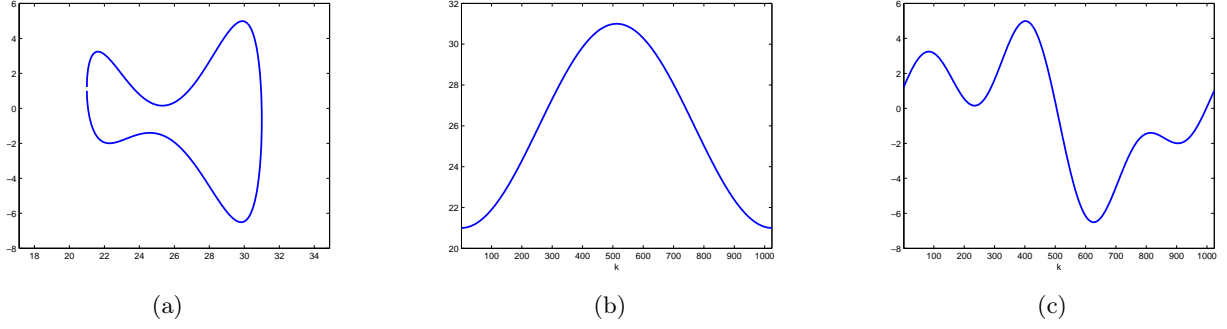


Figure 3: (a) Plot of the mean pattern  $\mathbf{f} = (\mathbf{f}^{(1)}, \mathbf{f}^{(2)})$  used in the simulations with  $k = 1024$ . (b) The first coordinates  $\mathbf{f}^{(1)}$ . (c) The second coordinates  $\mathbf{f}^{(2)}$ .

This curve is plotted in Figure 3. The deformation parameters  $(a_j^*, \alpha_j^*, b_j^*)$ ,  $j = 1, \dots, J$ , are i.i.d uniform random variables taking their values in  $\Theta = [-\frac{1}{4}, \frac{1}{4}] \times [-\frac{1}{2}, \frac{1}{2}] \times [-1, 1]^2$ . The law of the deformation parameters is supposed to be unknown *a priori* and the minimization is performed on the constraint set

$$\Theta_0 = \left\{ (\mathbf{a}, \boldsymbol{\alpha}, \mathbf{b}) \in [-1, 1]^J \times [-1, 1]^J \times [-5, 5]^{2J}, \sum_{j=1}^J a_j = 0, \sum_{j=1}^J \alpha_j = 0 \text{ and } \sum_{j=1}^J b_j = 0 \right\}.$$

Recall our notations: the error term is denoted by  $\boldsymbol{\zeta} = (\boldsymbol{\zeta}^{(1)}, \boldsymbol{\zeta}^{(2)}) \in \mathbb{R}^{k \times 2}$  and the vectorized version of  $\boldsymbol{\zeta}$  is denoted by  $\tilde{\boldsymbol{\zeta}} \in \mathbb{R}^{2k}$ . The simulations were run with three different kinds of noise.

*White noise* : the random variable  $\tilde{\boldsymbol{\zeta}} = (\zeta_1^{(1)}, \zeta_2^{(1)}, \dots, \zeta_1^{(k)}, \zeta_2^{(k)})' \in \mathbb{R}^{2k}$  is a centered Gaussian vector of variance

$$\boldsymbol{\Sigma}_1 = 4Id_{2k}.$$

We have  $\gamma_{\max}(k) = 4$  and this correspond to an isotropic Gaussian noise as in [Kent & Mardia, 1997, Le, 1998], see Figure 4.

*Weakly correlated noise* : the random variable  $\tilde{\boldsymbol{\zeta}}$  is a centered Gaussian vector of variance  $\boldsymbol{\Sigma}_2$  with

$$\boldsymbol{\Sigma}_2^{\frac{1}{2}} = Id_2 \otimes \left[ \frac{1}{2} \exp \left( -\frac{|\ell - \ell'|^2}{100} \right) \right]_{\ell, \ell'=1}^k$$

Hence,  $\boldsymbol{\Sigma}_2$  is a Toeplitz matrix and it follows from classical matrix theory, see e.g. [Horn & Johnson, 1990], that  $\gamma_{\max}(k)$  is bounded (here  $\gamma_{\max} \leq 80$ ). See Figure 5.

*Highly correlated noise* : the random variable  $\tilde{\boldsymbol{\zeta}}$  is a centered Gaussian vector of variance

$$\boldsymbol{\Sigma}_3 = Id_2 \otimes P \text{Diag} \left( \frac{\ell^2}{2k}, \ell = 1, \dots, k \right) P'$$

where  $P$  is an arbitrary matrix in  $\mathbf{SO}(k)$ . Hence, in this case  $\gamma_{\max}(k) = \frac{k}{4}$  and the level of noise increase with  $k$ , see Figure 6.

## 6.2 Description of the procedure

The estimation procedure follows the guidelines described in Section 4. We are testing the effect of the number  $J$  of observations and the number  $k$  of landmarks on the estimation of the parameters of interest of model (1.1). All the simulations are performed with  $J = 10, 100, 500$  and  $k = 20, 50, 100, 1000, 3000$



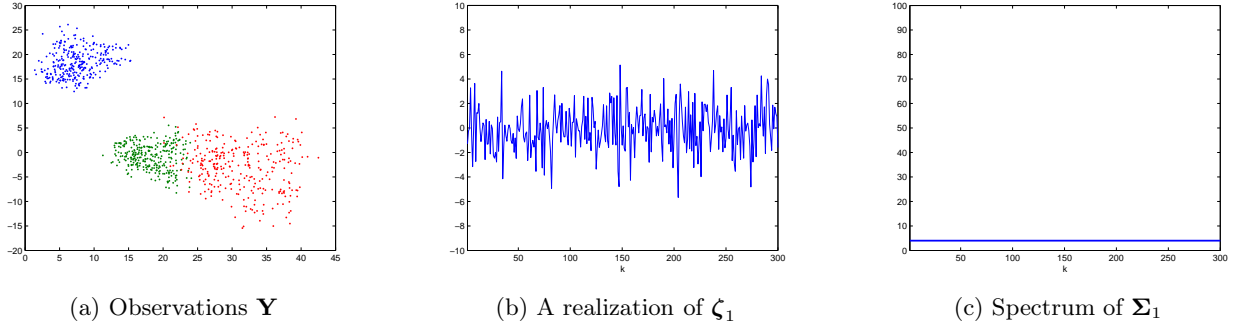


Figure 4: Example of data generated by model (1.1) with white noise.

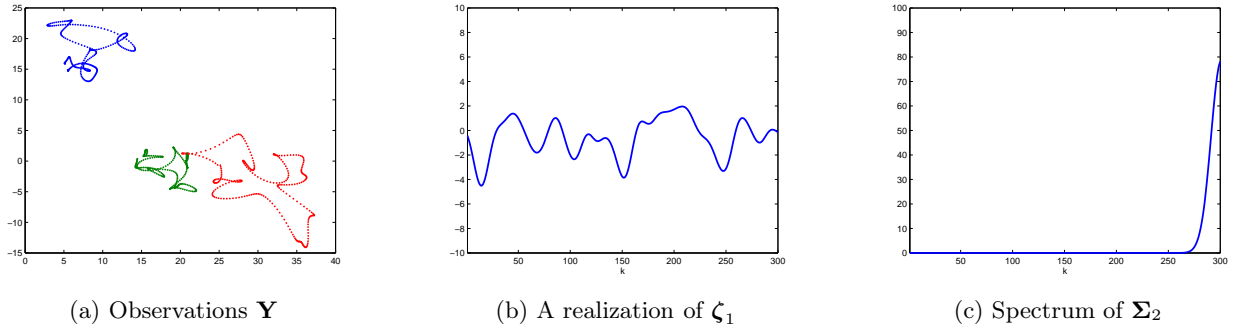


Figure 5: Example of data generated by model (1.1) with weakly correlated noise.

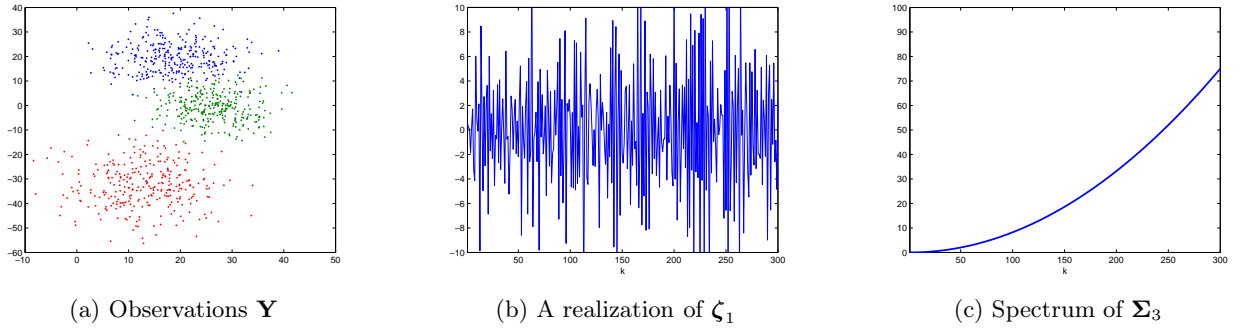


Figure 6: Example of observations generated by model (1.1) with the highly correlated noise.

and for each combination of these two factors the simulations are performed with  $M = 30$  repetitions of model (1.1).

Moreover, estimations are done without and with the pre-smoothing step. In the former case we have  $\lambda(k) = \frac{k}{2}$ , that is, there is no reduction of the dimension. In the latter case, the smoothing parameter  $\lambda(k)$  is fixed manually to ensure a proper reconstruction of the mean pattern  $f$ . Note that we need  $\lambda > 7$  to get correct results and we took  $\lambda_{20} = \lambda_{50} = \lambda_{100} = 7$ ,  $\lambda_{1000} = 11$  and  $\lambda_{3000} = 25$ .

We use a quasi-Newton trust-region based algorithm to solve the optimization problems (4.3) and (4.4). The formula for the gradient is given in (B.1). All the computations are performed with *Matlab*.

### 6.3 Results: estimation of the mean pattern

For each of the 30 repetitions of model (1.1) with the possible values of  $k$  and  $J$ , we compute the quantities  $\frac{1}{k} \|\hat{\mathbf{f}}^\lambda - \mathbf{f}_{\Theta_0}^*\|_{\mathbb{R}^{k \times 2}}^2$  where  $\hat{\mathbf{f}}^\lambda$  corresponds to the smoothed Procrustes mean of the observations defined in (1.8) and,  $\frac{1}{k} \|\hat{\mathbf{f}} - \mathbf{f}_{\Theta_0}^*\|_{\mathbb{R}^{k \times 2}}^2$  where  $\hat{\mathbf{f}}$  is the (non smoothed) Procrustes mean of the data. Recall that  $\mathbf{f}_{\Theta_0}$  is defined by formula (5.3).

Boxplots of the results are given in Figures 7, 8 and 9 for the different kinds of error terms described in Section 6.1. In the figures, the abscissa represents the different values of the number  $k$  of landmarks and boxplots in red correspond to  $J = 10$  observations, in green to  $J = 100$  observations and in blue to  $J = 500$  observations.

The estimation of the mean pattern with the white noise error term is given by Figure 7. In Figure 7a, for a fixed  $k$ , the non-smoothed version  $\frac{1}{k} \|\hat{\mathbf{f}} - \mathbf{f}^*\|_{\mathbb{R}^{k \times 2}}^2$  decreases when  $J$  increases. Moreover, the values of  $\frac{1}{k} \|\hat{\mathbf{f}} - \mathbf{f}^*\|_{\mathbb{R}^{k \times 2}}^2$  remain stable when  $J$  remains fixed and  $k$  increases. Recall that this framework corresponds to the isotropic Gaussian noise described in [Kent & Mardia, 1997]. The simulations seem to confirm their conclusions and show that in this framework the dimension  $k$  is not preponderant. In Figure 7b, the smoothed version  $\frac{1}{k} \|\hat{\mathbf{f}}^\lambda - \mathbf{f}\|_{\mathbb{R}^{k \times 2}}^2$  decreases when  $J$  and  $k$  increase. The main difference with the non-smoothed estimation is the convergence to 0 of  $\frac{1}{k} \|\hat{\mathbf{f}}^\lambda - \mathbf{f}\|_{\mathbb{R}^{k \times 2}}^2$  when  $J$  remains fixed and  $k$  increases.

In Figure 8, the results of the estimation of the mean pattern are plotted for the weakly correlated noise term. Figure 8a shows us a similar behavior of the non-smoothed Procrustes mean but with non-decreasing values of  $\frac{1}{k} \|\hat{\mathbf{f}}^\lambda - \mathbf{f}^*\|_{\mathbb{R}^{k \times 2}}^2$  when  $k$  increases and  $J$  remains fixed. In Figure 8b, the smoothed Procrustes mean converges as  $k$  goes to infinity and the bigger  $J$  is the faster the convergence is.

The results of the estimations of the mean pattern with the highly correlated noise are presented Figure 9. The results that appear in Figure 9a are quite different compared to those presented Figures 7a and 8a. The estimation seems to be worst when  $k$  increases and  $J$  remains fixed. The reason is that the level of noise, measured by  $\gamma_{\max}(k)$  is increasing with  $k$ . The smoothing step is efficient and the estimations presented Figure 9b have a similar behavior to those given in Figures 7b and 8b.

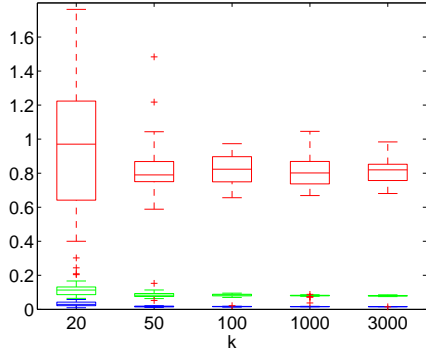
## A Proofs

### A.1 Proof of Lemma 4.1

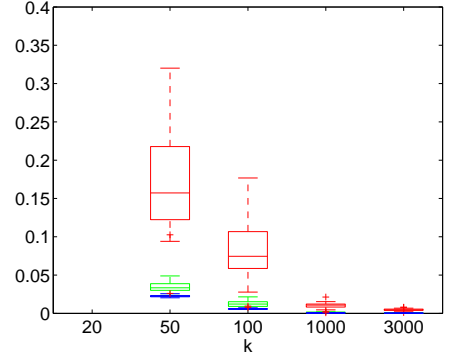
Using the decomposition (4.2), we obtain the following identity

$$M^\lambda(\mathbf{a}, \boldsymbol{\alpha}, \mathbf{b}) = M_0^\lambda(\mathbf{a}, \boldsymbol{\alpha}) + \bar{M}(\mathbf{a}, \boldsymbol{\alpha}, \mathbf{b}). \quad (\text{A.1})$$

Note that we have used the fact that the subspaces  $\mathcal{V}_0^\lambda$  and  $\bar{\mathcal{V}}$  are orthogonal and that  $\bar{A}\mathbf{1}_k \otimes b_j = \mathbf{1}_k \otimes b_j$ , for all  $b_j \in \mathbb{R}^2$ . Let us also introduce the notation  $\bar{\mathbf{Y}}_j = \frac{1}{k} \sum_{\ell=1}^k [\mathbf{Y}_j]_\ell \in \mathbb{R}^2$ , that is  $\bar{A}\mathbf{Y}_j = \mathbf{1}_k \otimes \bar{\mathbf{Y}}_j \in \mathbb{R}^{k \times 2}$  is a degenerated configuration, and  $\bar{\mathbf{Y}} = \frac{1}{J} \sum_{j=1}^J \bar{\mathbf{Y}}_j \in \mathbb{R}^2$ . For a fixed  $(\mathbf{a}, \boldsymbol{\alpha}) \in \mathbb{R}^J \times [-\pi, \pi]^J$ , the functional  $\mathbf{b} \mapsto \bar{M}(\mathbf{a}, \boldsymbol{\alpha}, \mathbf{b})$  vanishes if and only if there exists a  $b_0 \in \mathbb{R}^2$  such that  $e^{-a_j}(\bar{A}\mathbf{Y}_j - \mathbf{1}_k \otimes b_j)R_{-\alpha_j} = \mathbf{1}_k \otimes b_0$  for all  $j = 1, \dots, J$ . Therefore, for this fixed  $(\mathbf{a}, \boldsymbol{\alpha})$ , there is a unique point  $\mathbf{b} = \mathbf{b}(\mathbf{a}, \boldsymbol{\alpha}) := (\bar{\mathbf{Y}}_1 - e^{a_1} \bar{\mathbf{Y}}(e^{\mathbf{a}} R_{\boldsymbol{\alpha}})^{-1} R_{\alpha_1}, \dots, \bar{\mathbf{Y}}_J - e^{a_J} \bar{\mathbf{Y}}(e^{\mathbf{a}} R_{\boldsymbol{\alpha}})^{-1} R_{\alpha_J})$  with

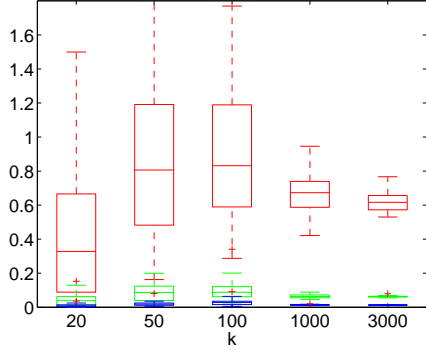


(a) Boxplot of  $\frac{1}{k} \|\hat{\mathbf{f}} - \mathbf{f}^*\|_{\mathbb{R}^{k \times 2}}^2$  (estimation without smoothing).

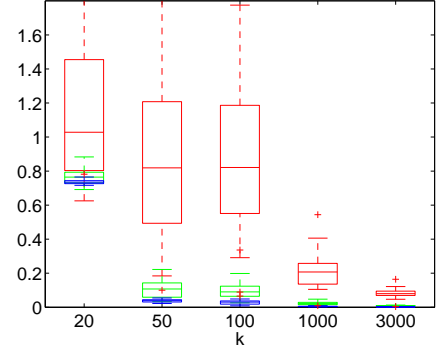


(b) Boxplot of  $\frac{1}{k} \|\hat{\mathbf{f}}^\lambda - \mathbf{f}^*\|_{\mathbb{R}^{k \times 2}}^2$  (estimation with smoothing). The missing boxplots for  $k = 20$  belong to the range  $[0.7, 1.9]$ .

Figure 7: Estimation of the mean pattern with the white noise. Boxplot in red correspond to  $J = 10$ , in green to  $J = 100$  and in blue to  $J = 500$ .

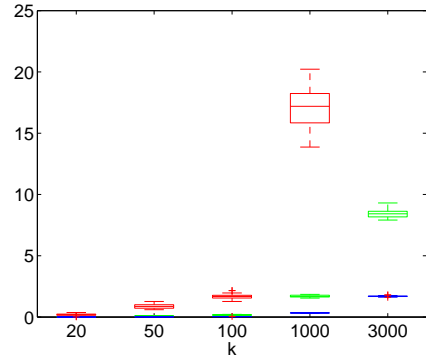


(a) Boxplot of  $\frac{1}{k} \|\hat{\mathbf{f}} - \mathbf{f}^*\|_{\mathbb{R}^{k \times 2}}^2$  (estimation without smoothing).

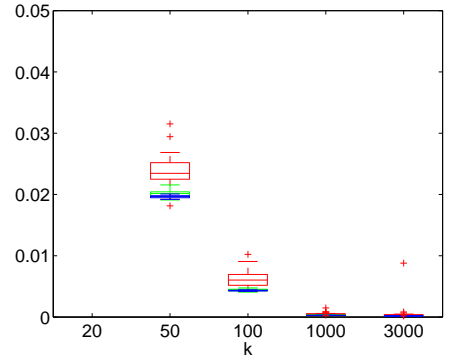


(b) Boxplot of  $\frac{1}{k} \|\hat{\mathbf{f}}^\lambda - \mathbf{f}^*\|_{\mathbb{R}^{k \times 2}}^2$  (estimation with smoothing).

Figure 8: Estimation of the mean pattern with the weakly correlated noise. Boxplot in red correspond to  $J = 10$ , in green to  $J = 100$  and in blue to  $J = 500$ .



(a) Boxplot of  $\frac{1}{k} \|\hat{\mathbf{f}} - \mathbf{f}^*\|_{\mathbb{R}^{k \times 2}}^2$  (estimation without smoothing). The missing red boxplot for  $k = 20$  and  $J = 10$  belongs to the range  $[70, 110]$ .



(b) Boxplot of  $\frac{1}{k} \|\hat{\mathbf{f}}^\lambda - \mathbf{f}^*\|_{\mathbb{R}^{k \times 2}}^2$  (estimation with smoothing). The missing boxplots for  $k = 20$  belong to the range  $[0.7, 1.9]$ .

Figure 9: Estimation of the mean pattern with the highly correlated noise. Boxplot in red correspond to  $J = 10$ , in green to  $J = 100$  and in blue to  $J = 500$ .

$\overline{e^a R_\alpha} = \frac{1}{J} \sum_{j=1}^J e^{a_j} R_{\alpha_j} \in \mathbb{R}^{2 \times 2}$  and which satisfies,

$$\mathbf{b}(\mathbf{a}, \boldsymbol{\alpha}) = \underset{\mathbf{b} \in \Theta_0^b}{\operatorname{argmin}} \bar{M}(\mathbf{a}, \boldsymbol{\alpha}, \mathbf{b})$$

Thence, thanks to the decomposition (A.1) and the fact that  $\bar{M}(\mathbf{a}, \boldsymbol{\alpha}, \mathbf{b}(\mathbf{a}, \boldsymbol{\alpha})) = 0$ , we have

$$\underset{(\mathbf{a}, \boldsymbol{\alpha}, \mathbf{b}) \in \Theta_0}{\operatorname{argmin}} M^\lambda(\mathbf{a}, \boldsymbol{\alpha}, \mathbf{b}) = \left( \underset{(\mathbf{a}, \boldsymbol{\alpha}) \in \Theta_0^{a, \alpha}}{\operatorname{argmin}} M_0^\lambda(\mathbf{a}, \boldsymbol{\alpha}), \mathbf{b} \left( \underset{(\mathbf{a}, \boldsymbol{\alpha}) \in \Theta_0^{a, \alpha}}{\operatorname{argmin}} M_0^\lambda(\mathbf{a}, \boldsymbol{\alpha}) \right) \right),$$

and the claim is proved.  $\square$

## A.2 Proof of Theorem 5.1

For all  $(\hat{\mathbf{a}}^\lambda, \hat{\boldsymbol{\alpha}}^\lambda) \in \Theta_0^{a, \alpha}$  and  $(\mathbf{a}^*, \boldsymbol{\alpha}^*) \in [-A, A] \times [-\mathcal{A}, \mathcal{A}]$ , we have the following inequality

$$\frac{1}{J} \|(\hat{\mathbf{a}}^\lambda, \hat{\boldsymbol{\alpha}}^\lambda) - (\mathbf{a}^*, \boldsymbol{\alpha}^*)\|_{\mathbb{R}^{2J}}^2 \leq \frac{2}{J} \|(\hat{\mathbf{a}}^\lambda, \hat{\boldsymbol{\alpha}}^\lambda) - (\mathbf{a}_{\Theta_0}^*, \boldsymbol{\alpha}_{\Theta_0}^*)\|_{\mathbb{R}^{2J}}^2 + \frac{2}{J} \|(\mathbf{a}_{\Theta_0}^*, \boldsymbol{\alpha}_{\Theta_0}^*) - (\mathbf{a}^*, \boldsymbol{\alpha}^*)\|_{\mathbb{R}^{2J}}^2.$$

The proof of Theorem 5.1 is a direct consequence of Proposition A.1 and Lemma A.2 below which control the convergence in probability of the two terms in the right hand side in the preceding inequality.

**Proposition A.1.** *Consider model (1.1) and suppose that Assumptions 1 and 2 hold and that Assumption 3 is verified with  $A, \mathcal{A} < 0.1$ . If  $\lambda = \lambda(k) = k^{\frac{1}{2s+1}}$  then there exists a constant  $C(L, s, A, \mathcal{A}, \mathbf{f}_0)$  such that for all  $x > 0$*

$$\mathbb{P} \left( \frac{1}{J} \|(\hat{\mathbf{a}}^\lambda, \hat{\boldsymbol{\alpha}}^\lambda) - (\mathbf{a}_{\Theta_0}^*, \boldsymbol{\alpha}_{\Theta_0}^*)\|_{\mathbb{R}^{2J}}^2 \geq C(L, s, A, \mathcal{A}, \mathbf{f}_0) \left( F(k^{-\frac{2s}{2s+1}}) + F(V_1(k, J, x)) \right) \right) \leq 2e^{-x},$$

where  $V_1(k, J, x) = 3\gamma_{\max}(k)k^{-\frac{2s}{2s+1}} (1 + \sqrt{2\frac{x}{Jk}} + 2\frac{x}{Jk})$  and  $F : \mathbb{R}^+ \rightarrow \mathbb{R}$ , with  $F(u) = u + \sqrt{u}$ .

The proof of Proposition A.1 is postponed to Section A.5. The following lemma is a direct consequence of Bernstein's inequality for bounded random variables, see e.g. Proposition 2.9 in [Massart, 2007].

**Lemma A.2.** *Suppose that Assumption 3 holds and that the random variables  $(a_j^*, \alpha_j^*)$ ,  $j = 1, \dots, J$  have zero expectation in  $[-A, A] \times [-\mathcal{A}, \mathcal{A}]$ . Then, for any  $x > 0$ , we have*

$$\mathbb{P} \left( \frac{1}{J} \|(\mathbf{a}_{\Theta_0}^*, \boldsymbol{\alpha}_{\Theta_0}^*) - (\mathbf{a}^*, \boldsymbol{\alpha}^*)\|_{\mathbb{R}^{2J}}^2 \geq C(A, \mathcal{A}) \left( \sqrt{\frac{2x}{J}} + \frac{x}{3J} \right)^2 \right) \leq 4e^{-x},$$

where  $C(A, \mathcal{A}) = 4 \max\{A^2, \mathcal{A}^2\}$ .  $\square$

## A.3 Proof of Theorem 5.2

The proof of Theorem 5.2 follows the same guideline as the proof of Theorem 5.1. Consider the inequality

$$\frac{1}{J} \|\hat{\mathbf{b}}^\lambda - \mathbf{b}^*\|_{\mathbb{R}^{2J}}^2 \leq \frac{2}{J} \|\hat{\mathbf{b}}^\lambda - \mathbf{b}_{\Theta_0}^*\|_{\mathbb{R}^{2J}}^2 + \frac{2}{J} \|\mathbf{b}_{\Theta_0}^* - \mathbf{b}^*\|_{\mathbb{R}^{2J}}^2.$$

Theorem 5.2 is now a direct consequence of Proposition A.3 and Lemma A.4.

**Proposition A.3.** *Under the hypothesis of Proposition A.1, there exists a constant  $C(L, s, A, \mathcal{A}, B, \mathbf{f}) > 0$  such that for all  $x > 0$ ,*

$$\mathbb{P} \left( \frac{2}{J} \|\hat{\mathbf{b}}^\lambda - \mathbf{b}_{\Theta_0}^*\|_{\mathbb{R}^{2J}}^2 \geq C(L, s, A, \mathcal{A}, B, \mathbf{f}) \left( F(k^{-\frac{2s}{2s+1}}) + F(V_1(k, J, x)) + V_2(k, J, x) \right) \right) \leq 5e^{-x},$$

where  $V_1(k, J, x) = 3\gamma_{\max}(k)k^{-\frac{2s}{2s+1}} (1 + \sqrt{2\frac{x}{Jk}} + 2\frac{x}{Jk})$ ,  $V_2(k, J, x) = \frac{4}{k}\gamma_{\max}(k)(1 + \sqrt{2\frac{x}{J}} + \frac{x}{J})$  and  $F(u) = u + \sqrt{u}$ ,  $u \geq 0$ .

The proof of Proposition A.3 is postponed to Section A.6.

**Lemma A.4.** *Suppose that Assumption 3 holds with  $\mathcal{A} < \frac{\pi}{4}$  and that the random variables  $(a_j^*, \alpha_j^*, b_j^*)$   $j = 1, \dots, J$  have zero expectation in  $\Theta \subset \mathbb{R}^{4J}$ . For any  $x > 0$ , we have*

$$\mathbb{P}\left(\frac{2}{J}\|\mathbf{b}_{\Theta_0}^* - \mathbf{b}^*\|_{\mathbb{R}^{2J}}^2 \geq C(A, \mathcal{A}, B)\left(\sqrt{\frac{2x}{J}} + \frac{x}{3J}\right)^2\right) \leq 4e^{-x},$$

where  $C(A, \mathcal{A}, B) = 8B^2e^{4A}(\cos \mathcal{A} - \sin \mathcal{A})^{-2}$ .

*Proof.* This result is a consequence of the Bernstein's inequality for bounded random variable. We have

$$\|\mathbf{b}_{\Theta_0}^* - \mathbf{b}^*\|_{\mathbb{R}^{2J}} = \sqrt{\sum_{j=1}^J \|e^{a_j^*} \bar{\mathbf{b}}^* (\overline{e^{\mathbf{a}} R_{\alpha}})^{-1} R_{\alpha_j^*}\|_{\mathbb{R}^2}^2} \leq e^A \sqrt{J} \|\bar{\mathbf{b}}^* (\overline{e^{\mathbf{a}} R_{\alpha}})^{-1}\|_{\mathbb{R}^2},$$

where  $\overline{e^{\mathbf{a}} R_{\alpha}} = \frac{1}{J} \sum_{j=1}^J e^{a_j^*} R_{\alpha_j^*}$  is an invertible  $2 \times 2$  matrix whose smallest eigenvalue is greater than  $e^{-A}(\cos \mathcal{A} - \sin \mathcal{A}) > 0$  as  $\mathcal{A} < \frac{\pi}{4}$ . To see this, remark that the eigenvalues of  $\overline{e^{\mathbf{a}} R_{\alpha}}$  are  $\frac{1}{J} \sum_{j=1}^J e^{a_j^*} (\cos \alpha_j^* \pm i \sin \alpha_j^*)$  and we have  $|\frac{1}{J} \sum_{j=1}^J e^{a_j^*} (\cos \alpha_j^* \pm i \sin \alpha_j^*)| \geq e^{-A}(\cos \mathcal{A} - \sin \mathcal{A}) > 0$ . We now have

$$\frac{1}{\sqrt{J}} \|\mathbf{b}_{\Theta_0}^* - \mathbf{b}^*\|_{\mathbb{R}^{2J}} \leq C(A, \mathcal{A}) \|\bar{\mathbf{b}}^*\|_{\mathbb{R}^2},$$

where  $C(A, \mathcal{A}) = e^{2A}(\cos \mathcal{A} - \sin \mathcal{A})^{-1}$ . Finally, for all  $u > 0$  we have  $\mathbb{P}(\frac{1}{\sqrt{J}} \|\mathbf{b}_{\Theta_0}^* - \mathbf{b}^*\|_{\mathbb{R}^{2J}} \geq u) \leq \mathbb{P}(C(A, \mathcal{A}) \|\bar{\mathbf{b}}^*\|_{\mathbb{R}^2} \geq u)$  and a Bernstein type inequality (see e.g. Proposition 2.9 in [Massart, 2007]) gives us  $\mathbb{P}(\|\bar{\mathbf{b}}^*\|_{\mathbb{R}^2} \geq 2B(\sqrt{\frac{2x}{J}} + \frac{x}{3J})) \leq 4e^{-x}$  which yields

$$\mathbb{P}\left(\frac{1}{J}\|\mathbf{b}_{\Theta_0}^* - \mathbf{b}^*\|_{\mathbb{R}^{2J}}^2 \geq C(A, \mathcal{A}, B)\left(\sqrt{\frac{2x}{J}} + \frac{x}{3J}\right)^2\right) \leq 4e^{-x},$$

where  $C(A, \mathcal{A}, B) = 4B^2e^{4A}(\cos \mathcal{A} - \sin \mathcal{A})^{-2}$ . □

#### A.4 Proof of Theorem 5.3

Recall the notations introduced Section 2.1: for  $\hat{g}_j^\lambda = (\hat{a}_j^\lambda, \hat{\alpha}_j^\lambda, \hat{b}_j^\lambda)$  we have  $\hat{g}_j^\lambda \cdot \mathbf{f} = e^{\hat{a}_j^\lambda} \mathbf{f} R_{\hat{\alpha}_j^\lambda} + \hat{b}_j^\lambda$ . Then, we have

$$\begin{aligned} \frac{1}{k} \|\hat{\mathbf{f}}^\lambda - \mathbf{f}\|_{\mathbb{R}^{k \times 2}}^2 &= \frac{1}{k} \left\| \frac{1}{J} \sum_{j=1}^J (\hat{g}_j^\lambda)^{-1} \cdot (\mathbf{A}^\lambda \mathbf{Y}_j) - \mathbf{f} \right\|_{\mathbb{R}^{k \times 2}}^2 \\ &\leq \frac{2}{kJ} \sum_{j=1}^J \left\| \mathbf{A}^\lambda \left( (\hat{g}_j^\lambda)^{-1} \cdot \mathbf{Y}_j - \mathbf{f} \right) \right\|_{\mathbb{R}^{k \times 2}}^2 + \frac{2}{k} \left\| \mathbf{A}^\lambda \mathbf{f} - \mathbf{f} \right\|_{\mathbb{R}^{k \times 2}}^2 \\ &\leq \underbrace{\frac{4}{kJ} \sum_{j=1}^J \left\| \mathbf{A}^\lambda \left( (\hat{g}_j^\lambda)^{-1} \cdot \mathbf{g}_j^* \cdot \mathbf{f} - \mathbf{f} \right) \right\|_{\mathbb{R}^{k \times 2}}^2}_{\mathbf{V}} + \underbrace{\left\| e^{a_j^* - \hat{a}_j^\lambda} \mathbf{A}^\lambda \zeta_j R_{\alpha_j^* - \hat{\alpha}_j^\lambda} \right\|_{\mathbb{R}^{k \times 2}}^2 + \frac{2}{k} \left\| \mathbf{A}^\lambda \mathbf{f} - \mathbf{f} \right\|_{\mathbb{R}^{k \times 2}}^2}_{\mathbf{B}}. \end{aligned}$$

The rest of the proof is devoted to control the terms  $\mathbf{B}$  and  $\mathbf{V}$ . The term  $\mathbf{B}$  is controlled by the bias of the low pass filter. According to Lemma B.1, if Assumption 1 holds and by choosing the optimal frequency cutoff  $\lambda = \lambda(k) = k^{\frac{1}{2s+1}}$ , there exists a constant  $C(L, s) > 0$  such that

$$\mathbf{B} = \frac{2}{k} \|\mathbf{A}^\lambda \mathbf{f} - \mathbf{f}\|_{\mathbb{R}^{k \times 2}}^2 \leq C(L, s) k^{-\frac{2s}{2s+1}}. \quad (\text{A.2})$$

The term  $\mathbf{V}$  contains two expressions. To bound the first one we use Bessel's inequality and Lemma B.4. More precisely, we have for all  $j = 1, \dots, J$

$$\begin{aligned} \frac{1}{k} \|\mathbf{A}^\lambda ((g_j^\lambda)^{-1} \cdot g_j^* \cdot \mathbf{f} - \mathbf{f})\|_{\mathbb{R}^{k \times 2}}^2 &\leq \frac{1}{k} \|(g_j^\lambda)^{-1} \cdot g_j^* \cdot \mathbf{f} - \mathbf{f}\|_{\mathbb{R}^{k \times 2}}^2 \\ &\leq C(A, \mathbf{f}) \|(a_j^* - \hat{a}_j^\lambda, \alpha_j^* - \hat{\alpha}_j^\lambda, e^{\hat{a}_j^\lambda} (b_j^* - \hat{b}_j^\lambda) R_{-\hat{\alpha}_j^\lambda})\|_{\mathbb{R}^4}^2 \\ &\leq C(A, \mathbf{f}) \|(a_j^* - \hat{a}_j^\lambda, \alpha_j^* - \hat{\alpha}_j^\lambda)\|_{\mathbb{R}^2}^2 + C(A, \mathbf{f}) \|b_j^* - \hat{b}_j^\lambda\|_{\mathbb{R}^2}^2 \end{aligned}$$

We can now use Theorem 5.1 and 5.2 to derive the following concentration inequality,

$$\begin{aligned} \mathbb{P} \left( \frac{4}{kJ} \sum_{j=1}^J \|\mathbf{A}^\lambda ((g_j^\lambda)^{-1} \cdot g_j^* \cdot \mathbf{f} - \mathbf{f})\|_{\mathbb{R}^{k \times 2}}^2 \right. \\ \left. \geq C(L, s, A, \mathcal{A}, B, \mathbf{f}) (A_1(k, J, x) + A_3(k, J, x) + A_2(J, x)) \right) \leq 13e^{-x} \end{aligned} \quad (\text{A.3})$$

where  $C(L, s, A, \mathcal{A}, B, \mathbf{f}) > 0$  is a constant independent of  $k$  and  $J$  and  $A_1, A_2, A_3$  are defined in the statement of Theorem 5.1 and 5.2. The second term contained in  $\mathbf{V}$  is treated by equation (A.10) below. Hence, formulas (A.3) and (A.10) yield

$$\mathbb{P} \left( \mathbf{V} \geq C(L, s, A, \mathcal{A}, B, \mathbf{f}) (A_1(k, J, x) + A_3(k, J, x) + A_2(J, x)) \right) \leq 14e^{-x}, \quad (\text{A.4})$$

for some constant  $C(L, s, A, \mathcal{A}, B, \mathbf{f}) > 0$ . Putting together equations (A.2) and (A.4) gives

$$\mathbb{P} \left( \frac{1}{k} \|\hat{\mathbf{f}}^\lambda - \mathbf{f}\|_{\mathbb{R}^{k \times 2}}^2 \geq C(L, s, A, \mathcal{A}, B, \mathbf{f}) (A_1(k, J, x) + A_3(k, J, x) + A_2(J, x)) \right) \leq 14e^{-x}.$$

for some constant  $C(L, s, A, \mathcal{A}, B, \mathbf{f}) > 0$ . The proof of Theorem 5.3 is completed.  $\square$

## A.5 Proof of Proposition A.1

The mean pattern  $\mathbf{f}$  can be decomposed as  $\mathbf{f} = \bar{\mathbf{f}} + \mathbf{f}_0 \in \mathbf{1}_{k \times 2} \oplus \mathbf{1}_{k \times 2}^\perp$ . Then,  $\mathbf{f}_0$  is the centered version of  $\mathbf{f}$  and we can consider the criterion,

$$D_0(\mathbf{a}, \boldsymbol{\alpha}) = \frac{1}{Jk} \sum_{j=1}^J \left\| e^{a_j^* - a_j} \mathbf{f}_0 R_{\alpha_j^* - \alpha_j} - \frac{1}{J} \sum_{j'=1}^J e^{a_{j'}^* - a_{j'}} \mathbf{f}_0 R_{\alpha_{j'}^* - \alpha_{j'}} \right\|_{\mathbb{R}^{k \times 2}}^2.$$

We now have,

$$(\hat{\mathbf{a}}^\lambda, \hat{\boldsymbol{\alpha}}^\lambda) = \underset{(\mathbf{a}, \boldsymbol{\alpha}) \in \boldsymbol{\Theta}_0^{\mathbf{a}, \boldsymbol{\alpha}}}{\operatorname{argmin}} M_0^\lambda(\mathbf{a}, \boldsymbol{\alpha}) \quad \text{and} \quad (\mathbf{a}_{\boldsymbol{\Theta}_0}^*, \boldsymbol{\alpha}_{\boldsymbol{\Theta}_0}^*) = \underset{(\mathbf{a}, \boldsymbol{\alpha}) \in \boldsymbol{\Theta}_0^{\mathbf{a}, \boldsymbol{\alpha}}}{\operatorname{argmin}} D_0(\mathbf{a}, \boldsymbol{\alpha}).$$

Then, the convergence of  $(\hat{\mathbf{a}}^\lambda, \hat{\boldsymbol{\alpha}}^\lambda)$  to  $(\mathbf{a}_{\boldsymbol{\Theta}_0}^*, \boldsymbol{\alpha}_{\boldsymbol{\Theta}_0}^*)$  is guaranteed if  $(\mathbf{a}_{\boldsymbol{\Theta}_0}^*, \boldsymbol{\alpha}_{\boldsymbol{\Theta}_0}^*)$  is uniquely defined and if there is a uniform convergence in probability of  $M_0^\lambda$  to  $D_0$ , see e.g. [van der Vaart, 1998]. This is the aim of Lemmas A.5 and A.6 below.

**Lemma A.5.** *Let  $\mathbf{f}$  be a non-degenerated configuration in  $\mathbb{R}^{k \times 2}$ , i.e.  $\mathbf{f} \notin \mathbf{1}_{k \times 2}$ . Then, the argmin of  $D_0$  on  $\boldsymbol{\Theta}_0^{\mathbf{a}, \boldsymbol{\alpha}}$  is unique and denoted by  $(\mathbf{a}_{\boldsymbol{\Theta}_0}^*, \boldsymbol{\alpha}_{\boldsymbol{\Theta}_0}^*) = (a_j - \bar{\mathbf{a}}^*, \alpha_j - \bar{\boldsymbol{\alpha}}^*)$ , where  $\bar{\mathbf{a}}^* = \frac{1}{J} \sum_{j=1}^J a_j^*$ ,  $\bar{\boldsymbol{\alpha}}^* = \frac{1}{J} \sum_{j=1}^J \alpha_j^*$ .*

*Proof.* As  $\mathbf{f}$  is a non-degenerated configuration, we have  $\mathbf{f}_0 \neq 0$ . Thus, the stabilizer of  $\mathbf{f}$  is reduced to the identity, see Section 2.1. Then  $D_0(\mathbf{a}, \boldsymbol{\alpha}) = 0$  if and only if there exists  $(a_0, \alpha_0) \in \mathbb{R}^2$  such that  $(\mathbf{a}, \boldsymbol{\alpha}) = (\mathbf{a}^*, \boldsymbol{\alpha}^*) * (a_0, \alpha_0) = (a_1^* + a_0, \alpha_1^* + \alpha_0, \dots, a_J^* + a_0, \alpha_J^* + \alpha_0)$ . By choosing  $(a_0, \alpha_0) = (-\bar{\mathbf{a}}^*, -\bar{\boldsymbol{\alpha}}^*)$  we have  $\sum_{j=1}^J (a_j^*, \alpha_j^*) * (a_0, \alpha_0) = 0$ . That is  $(\mathbf{a}_{\boldsymbol{\Theta}_0}^*, \boldsymbol{\alpha}_{\boldsymbol{\Theta}_0}^*) = (\mathbf{a}^*, \boldsymbol{\alpha}^*) * (a_0, \alpha_0) \in \boldsymbol{\Theta}_0^{\mathbf{a}, \boldsymbol{\alpha}}$ .  $\square$

We now show the uniform convergence in probability,

**Lemma A.6.** *Suppose that Assumptions 1, 2 and 3 hold and let  $F : \mathbb{R} \rightarrow \mathbb{R}$ , with  $F(u) = u + \sqrt{u}$ . For any  $x > 0$  we have*

$$\mathbb{P}\left(\sup_{(\mathbf{a}, \boldsymbol{\alpha}) \in \Theta_0^{\mathbf{a}, \boldsymbol{\alpha}}} |M_0^\lambda(\mathbf{a}, \boldsymbol{\alpha}) - D_0(\mathbf{a}, \boldsymbol{\alpha})| \geq C(L, s, A, \mathbf{f}_0) \left(F(k^{-\frac{2s}{2s+1}}) + F(V(k, J, x))\right)\right) \leq 2e^{-x}$$

where  $C(L, s, A, \mathbf{f}_0) = e^{2A} \max\left\{\frac{2}{\sqrt{k}}\|\mathbf{f}_0\|_{\mathbb{R}^{k \times 2}}, \frac{\sqrt{2}e^A}{\sqrt{k}}\|\mathbf{f}_0\|_{\mathbb{R}^{k \times 2}}, 2\right\}$  and  $V(k, J, x) = 3\gamma_{\max}(k)k^{-\frac{2s}{2s+1}}(1 + F(\frac{2x}{Jk}))$ .

*Proof.* Let us write the following decomposition,

$$M_0^\lambda(\mathbf{a}, \boldsymbol{\alpha}) = \frac{1}{Jk} \sum_{j=1}^J \left\| e^{a_j^* - a_j} \mathbf{f}_0 R_{\alpha_j^* - \alpha_j} - \frac{1}{J} \sum_{j'=1}^J e^{a_{j'}^* - a_{j'}} \mathbf{f}_0 R_{\alpha_{j'}^* - \alpha_{j'}} \right\|_{\mathbb{R}^{k \times 2}}^2 \quad (\text{A.5})$$

$$+ \frac{1}{Jk} \sum_{j=1}^J \left\| e^{a_j^* - a_j} (A_0^\lambda \mathbf{f} - \mathbf{f}_0) R_{\alpha_j^* - \alpha_j} - \frac{1}{J} \sum_{j'=1}^J e^{a_{j'}^* - a_{j'}} (A_0^\lambda \mathbf{f} - \mathbf{f}_0) R_{\alpha_{j'}^* - \alpha_{j'}} \right\|_{\mathbb{R}^{k \times 2}}^2 \quad (\text{A.6})$$

$$+ \frac{2}{Jk} \sum_{j=1}^J \left\langle e^{a_j^* - a_j} \mathbf{f}_0 R_{\alpha_j^* - \alpha_j} - \frac{1}{J} \sum_{j'=1}^J e^{a_{j'}^* - a_{j'}} \mathbf{f}_0 R_{\alpha_{j'}^* - \alpha_{j'}}, \right. \\ \left. e^{a_j^* - a_j} (A_0^\lambda \mathbf{f} - \mathbf{f}_0) R_{\alpha_j^* - \alpha_j} - \frac{1}{J} \sum_{j'=1}^J e^{a_{j'}^* - a_{j'}} (A_0^\lambda \mathbf{f} - \mathbf{f}_0) R_{\alpha_{j'}^* - \alpha_{j'}} \right\rangle_{\mathbb{R}^{k \times 2}} \quad (\text{A.7})$$

$$+ \frac{1}{Jk} \sum_{j=1}^J \left\| e^{a_j^* - a_j} A_0^\lambda \boldsymbol{\zeta}_j R_{\alpha_j^* - \alpha_j} - \frac{1}{J} \sum_{j'=1}^J e^{a_{j'}^* - a_{j'}} A_0^\lambda \boldsymbol{\zeta}_{j'} R_{\alpha_{j'}^* - \alpha_{j'}} \right\|_{\mathbb{R}^{k \times 2}}^2 \quad (\text{A.8})$$

$$+ \frac{2}{Jk} \sum_{j=1}^J \left\langle e^{a_j^* - a_j} A_0^\lambda \mathbf{f} R_{\alpha_j^* - \alpha_j} - \frac{1}{J} \sum_{j'=1}^J e^{a_{j'}^* - a_{j'}} A_0^\lambda \mathbf{f} R_{\alpha_{j'}^* - \alpha_{j'}}, \right. \\ \left. e^{a_j^* - a_j} A_0^\lambda \boldsymbol{\zeta}_j R_{\alpha_j^* - \alpha_j} - \frac{1}{J} \sum_{j'=1}^J e^{a_{j'}^* - a_{j'}} A_0^\lambda \boldsymbol{\zeta}_{j'} R_{\alpha_{j'}^* - \alpha_{j'}} \right\rangle_{\mathbb{R}^{k \times 2}} \quad (\text{A.9})$$

Then, criterion  $M_0^\lambda$  is viewed as a perturbed version of the criterion  $D_0(\mathbf{a}, \boldsymbol{\alpha}) = (\text{A.5})$ ,

$$M_0^\lambda(\mathbf{a}, \boldsymbol{\alpha}) = D_0(\mathbf{a}, \boldsymbol{\alpha}) + \mathbf{B}_0^\lambda(\mathbf{a}, \boldsymbol{\alpha}) + \mathbf{V}_0^\lambda(\mathbf{a}, \boldsymbol{\alpha})$$

where the bias term is  $\mathbf{B}_0^\lambda(\mathbf{a}, \boldsymbol{\alpha}) = (\text{A.6}) + (\text{A.7})$  and the variance term is  $\mathbf{V}_0^\lambda(\mathbf{a}, \boldsymbol{\alpha}) = (\text{A.8}) + (\text{A.9})$ .

**The bias term.** We have for any  $(\mathbf{a}, \boldsymbol{\alpha}) \in \Theta_0^{\mathbf{a}, \boldsymbol{\alpha}}$ ,

$$(\text{A.6}) = \frac{1}{Jk} \sum_{j=1}^J \left\| e^{a_j^* - a_j} (A_0^\lambda \mathbf{f} - \mathbf{f}_0) R_{\alpha_j^* - \alpha_j} \right\|_{\mathbb{R}^{k \times 2}}^2 - \left\| \frac{1}{J} \sum_{j'=1}^J e^{a_{j'}^* - a_{j'}} (A_0^\lambda \mathbf{f} - \mathbf{f}_0) R_{\alpha_{j'}^* - \alpha_{j'}} \right\|_{\mathbb{R}^{k \times 2}}^2 \\ \leq e^{2A} \frac{1}{k} \left\| A_0^\lambda \mathbf{f} - \mathbf{f}_0 \right\|_{\mathbb{R}^{k \times 2}}^2.$$

A double application of Cauchy-Schwarz inequality implies that,

$$\begin{aligned}
|(A.7)| &\leq \frac{2}{Jk} \sum_{j=1}^J \left\| e^{a_j^* - a_j} \mathbf{f}_0 R_{\alpha_j^* - \alpha_j} - \frac{1}{J} \sum_{j'=1}^J e^{a_{j'}^* - a_{j'}} \mathbf{f}_0 R_{\alpha_{j'}^* - \alpha_{j'}} \right\|_{\mathbb{R}^{k \times 2}} \\
&\quad \left\| e^{a_j^* - a_j} (\mathbf{A}_0^\lambda \mathbf{f} - \mathbf{f}_0) R_{\alpha_j^* - \alpha_j} - \frac{1}{J} \sum_{j'=1}^J e^{a_{j'}^* - a_{j'}} (\mathbf{A}_0^\lambda \mathbf{f} - \mathbf{f}_0) R_{\alpha_{j'}^* - \alpha_{j'}} \right\|_{\mathbb{R}^{k \times 2}} \\
&\leq \frac{2}{k} \left( \frac{1}{J} \sum_{j=1}^J e^{2A} \|\mathbf{f}_0 R_{\alpha_j^* - \alpha_j}\|_{\mathbb{R}^{k \times 2}}^2 \right)^{\frac{1}{2}} \left( \frac{1}{J} \sum_{j=1}^J e^{2A} \|(\mathbf{A}_0^\lambda \mathbf{f} - \mathbf{f}_0) R_{\alpha_j^* - \alpha_j}\|_{\mathbb{R}^{k \times 2}}^2 \right)^{\frac{1}{2}} \\
&= 2e^{2A} \frac{1}{k} \|\mathbf{f}_0\|_{\mathbb{R}^{k \times 2}} \|\mathbf{A}_0^\lambda \mathbf{f} - \mathbf{f}_0\|_{\mathbb{R}^{k \times 2}}.
\end{aligned}$$

Finally, by using Lemma B.1 we have,

$$\sup_{(\mathbf{a}, \boldsymbol{\alpha}) \in \boldsymbol{\Theta}^{\mathbf{a}, \boldsymbol{\alpha}}} |\mathbf{B}_0^\lambda(\mathbf{a}, \boldsymbol{\alpha})| \leq C_1(L, s, A, \mathbf{f}) \left( k^{-\frac{2s}{2s+1}} + k^{-\frac{s}{2s+1}} \right).$$

where  $C_1(L, s, A, \mathbf{f}) = C(L, s) e^{2A} \max \left\{ 2 \frac{1}{\sqrt{k}} \|\mathbf{f}_0\|_{\mathbb{R}^{k \times 2}}, 1 \right\}$ .

**The variance term.** First, the term (A.9) is by the Cauchy-Schwarz inequality controlled by  $2e^{2A} \frac{1}{\sqrt{k}} \|\mathbf{f}_0\|_{\mathbb{R}^{k \times 2}} \sqrt{(A.8)}$ . The term (A.8) is bounded from above by  $\frac{1}{Jk} \sum_{j=1}^J e^{2A} \|A_0^\lambda \boldsymbol{\zeta}_j\|_{\mathbb{R}^{k \times 2}}^2$ . To derive an upper bound in probability, note that we have the following equality in law,

$$\sum_{j=1}^J \|A_0^\lambda \boldsymbol{\zeta}_j\|_{\mathbb{R}^{k \times 2}}^2 = \boldsymbol{\xi}' \mathbf{B} \boldsymbol{\xi},$$

with  $\mathbf{B} = [Id_J \otimes \boldsymbol{\Sigma}^{\frac{1}{2}}] [Id_{2J} \otimes (A_0^\lambda)' A_0^\lambda] [Id_J \otimes \boldsymbol{\Sigma}^{\frac{1}{2}}] \in \mathbb{R}^{2Jk \times 2Jk}$  and  $\boldsymbol{\xi} = (\xi_1, \dots, \xi_{2Jk})'$  is a centered Gaussian vector of variance  $Id_{2Jk}$ . We have  $\text{tr}(Id_J \otimes \boldsymbol{\Sigma}) \leq 2Jk\gamma_{\max}(k)$  and  $\text{tr}((A_0^\lambda)' A_0^\lambda) = \frac{2\lambda+1}{k}$ . Using a classical concentration inequality for quadratic form of multivariate Gaussian random variables, see e.g. [Laurent & Massart, 2000] Lemma 1, we have for all  $x > 0$ ,  $\mathbb{P}(\boldsymbol{\xi}' \mathbf{B} \boldsymbol{\xi} \geq 2Jk\gamma_{\max}(k) \frac{2\lambda+1}{k} + 2\gamma_{\max}(k) \frac{2\lambda+1}{k} \sqrt{x2Jk} + 4x\gamma_{\max}(k) \frac{2\lambda+1}{k}) \leq e^{-x}$ , which yields together with formula (A.2),

$$\mathbb{P} \left( \frac{1}{Jk} \sum_{j=1}^J e^{2A} \|A_0^\lambda \boldsymbol{\zeta}_j\|_{\mathbb{R}^{k \times 2}}^2 \geq 6e^{2A} \gamma_{\max}(k) k^{-\frac{2s}{2s+1}} \left( 1 + \sqrt{2 \frac{x}{Jk}} + 2 \frac{x}{Jk} \right) \right) \leq e^{-x}. \quad (A.10)$$

Hence we have,

$$\begin{aligned}
&\mathbb{P} \left( \sup_{(\mathbf{a}, \boldsymbol{\alpha}) \in \boldsymbol{\Theta}^{\mathbf{a}, \boldsymbol{\alpha}}} |\mathbf{V}_0^\lambda(\mathbf{a}, \boldsymbol{\alpha}) + \mathbf{B}_0^\lambda(\mathbf{a}, \boldsymbol{\alpha})| \right. \\
&\quad \left. \geq C(L, s, A, \mathbf{f}_0) \left( k^{-\frac{2s}{2s+1}} + k^{-\frac{s}{2s+1}} + V_1(k, J, x) + \sqrt{V_1(k, J, x)} \right) \right) \leq 2e^{-x},
\end{aligned}$$

where  $C(L, s, A, \mathbf{f}_0) = e^{2A} \max \left\{ \frac{2}{\sqrt{k}} \|\mathbf{f}_0\|_{\mathbb{R}^{k \times 2}}, \frac{1}{\sqrt{k}} \|\mathbf{f}_0\|_{\mathbb{R}^{k \times 2}} \sqrt{2e^A}, 2 \right\}$  and  $V_1(k, J, x) = 3 \frac{\gamma_{\max}(k)}{k^{\frac{2s}{2s+1}}} (1 + \sqrt{2 \frac{x}{Jk}} + 2 \frac{x}{Jk})$ .  $\square$

For a fixed  $J \in \mathbb{N}$ , the convergence of the M-estimator  $(\hat{\mathbf{a}}^\lambda, \hat{\boldsymbol{\alpha}}^\lambda)$  to  $(\mathbf{a}_{\boldsymbol{\Theta}_0}^*, \boldsymbol{\alpha}_{\boldsymbol{\Theta}_0}^*)$  when  $k \rightarrow \infty$  is guaranteed by Lemma A.5 and A.6, see e.g. [van der Vaart, 1998]. Nevertheless, we are able to give a rate of convergence and non-asymptotic bounds in  $k$  and  $J$  by using the classical inequality,

$$|D_0(\hat{\mathbf{a}}^\lambda, \hat{\boldsymbol{\alpha}}^\lambda) - D_0(\mathbf{a}_{\boldsymbol{\Theta}_0}^*, \boldsymbol{\alpha}_{\boldsymbol{\Theta}_0}^*)| \leq 2 \sup_{(\mathbf{a}, \boldsymbol{\alpha}) \in \boldsymbol{\Theta}_0^{\mathbf{a}, \boldsymbol{\alpha}}} |D_0(\mathbf{a}, \boldsymbol{\alpha}) - M_0^\lambda(\mathbf{a}, \boldsymbol{\alpha})|.$$

This, together with Lemma A.7 below will prove Proposition A.1.



**Lemma A.7.** Assume that  $A, \mathcal{A} < 0.1$ . There exists a constant  $C(A, \mathcal{A}) > 0$  independent of  $J$  such that for all  $(\mathbf{a}, \boldsymbol{\alpha}) \in \boldsymbol{\Theta}^{\mathbf{a}, \boldsymbol{\alpha}}$  we have

$$|D_0(\mathbf{a}, \boldsymbol{\alpha}) - D_0(\mathbf{a}_{\Theta_0}^*, \boldsymbol{\alpha}_{\Theta_0}^*)| \geq C(A, \mathcal{A}, \mathbf{f}_0) \frac{1}{J} \|(\mathbf{a} - \mathbf{a}_{\Theta_0}^*, \boldsymbol{\alpha} - \boldsymbol{\alpha}_{\Theta_0}^*)\|_{\mathbb{R}^{2J}}^2,$$

where  $C(A, \mathcal{A}, \mathbf{f}_0) = C(A, \mathcal{A}) \frac{1}{k} \|\mathbf{f}_0\|_{\mathbb{R}^{k \times 2}}^2$ .

*Proof.* By definition, given a  $(\mathbf{a}^*, \boldsymbol{\alpha}^*) \in [-A, A]^J \times [-\mathcal{A}, \mathcal{A}]^J$ , the point  $(\mathbf{a}_{\Theta_0}^*, \boldsymbol{\alpha}_{\Theta_0}^*)$  is the unique minimum of  $D_0$  on  $\boldsymbol{\Theta}_0^{\mathbf{a}, \boldsymbol{\alpha}} = [-A, A]^J \cap \mathbb{1}_J^\perp \times [-\mathcal{A}, \mathcal{A}]^J \cap \mathbb{1}_J^\perp$ . Then, for all  $(\mathbf{a}, \boldsymbol{\alpha}) \in \boldsymbol{\Theta}_0^{\mathbf{a}, \boldsymbol{\alpha}}$ , there exists a  $\mathbf{c} = \mathbf{c}(\mathbf{a}, \boldsymbol{\alpha}) \in \boldsymbol{\Theta}_0^{\mathbf{a}, \boldsymbol{\alpha}}$  such that the Taylor expansion of  $D_0$  at  $(\mathbf{a}_{\Theta_0}^*, \boldsymbol{\alpha}_{\Theta_0}^*)$  can be written,

$$\begin{aligned} D_0(\mathbf{a}, \boldsymbol{\alpha}) - D_0(\mathbf{a}_{\Theta_0}^*, \boldsymbol{\alpha}_{\Theta_0}^*) &= \frac{1}{2}(\mathbf{a} - \mathbf{a}_{\Theta_0}^*, \boldsymbol{\alpha} - \boldsymbol{\alpha}_{\Theta_0}^*)' [\nabla^2 D_0(\mathbf{a}_{\Theta_0}^*, \boldsymbol{\alpha}_{\Theta_0}^*)] (\mathbf{a} - \mathbf{a}_{\Theta_0}^*, \boldsymbol{\alpha} - \boldsymbol{\alpha}_{\Theta_0}^*) \\ &\quad + \frac{1}{6} [\nabla^3 D_0(\mathbf{c})] (\mathbf{a} - \mathbf{a}_{\Theta_0}^*, \boldsymbol{\alpha} - \boldsymbol{\alpha}_{\Theta_0}^*). \end{aligned}$$

Let  $\delta = \max\{A, \mathcal{A}\}$ . This, together with Lemma B.2 and B.3 imply that

$$\begin{aligned} D_0(\mathbf{a}, \boldsymbol{\alpha}) - D_0(\mathbf{a}_{\Theta_0}^*, \boldsymbol{\alpha}_{\Theta_0}^*) &\geq \frac{1}{2}(\mathbf{a} - \mathbf{a}_{\Theta_0}^*, \boldsymbol{\alpha} - \boldsymbol{\alpha}_{\Theta_0}^*)' [\nabla^2 D_0(\mathbf{a}_{\Theta_0}^*, \boldsymbol{\alpha}_{\Theta_0}^*)] (\mathbf{a} - \mathbf{a}_{\Theta_0}^*, \boldsymbol{\alpha} - \boldsymbol{\alpha}_{\Theta_0}^*) \\ &\quad - \delta \frac{40}{6} e^{2A} \|\mathbf{f}_0\|_{\mathbb{R}^{k \times 2}}^2 \frac{1}{Jk} \|(\mathbf{a} - \mathbf{a}_{\Theta_0}^*, \boldsymbol{\alpha} - \boldsymbol{\alpha}_{\Theta_0}^*)\|_{\mathbb{R}^{2J}}^2 \\ &\geq \|\mathbf{f}_0\|_{\mathbb{R}^{k \times 2}}^2 \frac{1}{Jk} \|(\mathbf{a} - \mathbf{a}_{\Theta_0}^*, \boldsymbol{\alpha} - \boldsymbol{\alpha}_{\Theta_0}^*)\|_{\mathbb{R}^{2J}}^2 \left( e^{-2A} - \delta \frac{40}{6} e^{2A} \right). \end{aligned}$$

Hence, one can choose  $\delta > 0$  sufficiently small such that  $(e^{-2A} - \delta \frac{40}{6} e^{2A})$  is strictly positive for all  $J$  and  $k$ . For example, we have  $(e^{-2\delta} - \delta \frac{40}{6} e^{2\delta}) > 0$ , if  $\delta < 0.1$ . Then, using such a  $\delta$  it follows that for all  $(\mathbf{a}, \boldsymbol{\alpha}) \in \boldsymbol{\Theta}_0^{\mathbf{a}, \boldsymbol{\alpha}}$ ,

$$|D_0(\mathbf{a}, \boldsymbol{\alpha}) - D_0(\mathbf{a}_{\Theta_0}^*, \boldsymbol{\alpha}_{\Theta_0}^*)| \geq C(A, \mathcal{A}) \frac{1}{k} \|\mathbf{f}_0\|_{\mathbb{R}^{k \times 2}}^2 \frac{1}{J} \|(\mathbf{a} - \mathbf{a}_{\Theta_0}^*, \boldsymbol{\alpha} - \boldsymbol{\alpha}_{\Theta_0}^*)\|_{\mathbb{R}^{2J}}^2. \quad \square$$

The proof of Proposition A.1 is almost done. Remark that Lemma A.7 ensures that for all  $u \geq 0$  we have  $\mathbb{P}(\frac{1}{J} \|(\hat{\mathbf{a}}^\lambda, \hat{\boldsymbol{\alpha}}^\lambda) - (\mathbf{a}_{\Theta_0}^*, \boldsymbol{\alpha}_{\Theta_0}^*)\|_{\mathbb{R}^{2J}}^2 \geq u) \leq \mathbb{P}(\frac{2}{C(A, \mathcal{A}, \mathbf{f}_0)} \sup_{(\mathbf{a}, \boldsymbol{\alpha}) \in \boldsymbol{\Theta}_0^{\mathbf{a}, \boldsymbol{\alpha}}} |M_0^\lambda(\mathbf{a}, \boldsymbol{\alpha}) - D_0(\mathbf{a}, \boldsymbol{\alpha})| \geq u)$ . Lemma A.6 ensures that there is a constant  $C(L, s, A, \mathcal{A}, \mathbf{f}_0)$  such that for all  $x > 0$ ,

$$\mathbb{P}\left(\frac{1}{J} \|(\hat{\mathbf{a}}^\lambda, \hat{\boldsymbol{\alpha}}^\lambda) - (\mathbf{a}_{\Theta_0}^*, \boldsymbol{\alpha}_{\Theta_0}^*)\|_{\mathbb{R}^{2J}}^2 \geq C(L, s, A, \mathcal{A}, \mathbf{f}_0) (F(k^{-\frac{2s}{2s+1}}) + F(V_1(k, J, x)))\right) \leq 2e^{-x}.$$

where  $V_1(k, J, x) = \gamma_{\max}(k) k^{-\frac{2s}{2s+1}} (1 + \sqrt{2\frac{x}{Jk}} + 2\frac{x}{Jk})$  and  $F: \mathbb{R} \rightarrow \mathbb{R}$ , with  $F(u) = u + \sqrt{u}$ .  $\square$

## A.6 Proof of Proposition A.3

First of all remark that, thanks to formulas (3.3) and (4.5), we have explicit expressions of  $\mathbf{b}_{\Theta_0}^* = \operatorname{argmin}_{\mathbf{b} \in \boldsymbol{\Theta}_0^b} \bar{D}(\mathbf{a}^*, \boldsymbol{\alpha}^*, \mathbf{b})$  and of  $\hat{\mathbf{b}}^\lambda = \operatorname{argmin}_{\mathbf{b} \in \boldsymbol{\Theta}_0^b} \bar{M}(\hat{\mathbf{a}}^\lambda, \hat{\boldsymbol{\alpha}}^\lambda, \mathbf{b})$ . Indeed, we have,

$$\mathbf{b}_{\Theta_0}^* = (b_1^* - e^{a_1^*} \bar{\mathbf{b}}^* (\overline{e^{a^*} R_{\alpha^*}})^{-1} R_{\alpha_1^*}, \dots, b_J^* - e^{a_J^*} \bar{\mathbf{b}}^* (\overline{e^{a^*} R_{\alpha^*}})^{-1} R_{\alpha_J^*}),$$

where  $\overline{e^{a^*} R_{\alpha^*}} = \frac{1}{J} \sum_{j=1}^J e^{a_j^*} R_{\alpha_j^*} \in \mathbb{R}^{2 \times 2}$ ,  $\bar{\mathbf{b}}^* = \frac{1}{J} \sum_{j=1}^J b_j^* \in \mathbb{R}^2$  and

$$\hat{\mathbf{b}}^\lambda = (\bar{\mathbf{Y}}_1 - e^{\hat{a}_1^\lambda} \bar{\mathbf{Y}} (\overline{e^{\hat{a}^\lambda} R_{\hat{\alpha}^\lambda}})^{-1} R_{\hat{\alpha}_1^\lambda}, \dots, \bar{\mathbf{Y}}_J - e^{\hat{a}_J^\lambda} \bar{\mathbf{Y}} (\overline{e^{\hat{a}^\lambda} R_{\hat{\alpha}^\lambda}})^{-1} R_{\hat{\alpha}_J^\lambda})$$

where  $\overline{e^{\hat{a}^\lambda} R_{\hat{\alpha}^\lambda}} = \frac{1}{J} \sum_{j=1}^J e^{\hat{a}_j^\lambda} R_{\hat{\alpha}_j^\lambda} \in \mathbb{R}^{2 \times 2}$ ,  $\bar{\mathbf{Y}}_j = \frac{1}{k} \sum_{\ell=1}^k [\mathbf{Y}_j]_\ell = e^{a_j^*} \bar{\mathbf{f}} R_{\alpha_j^*} + b_j^* + e^{a_j^*} \bar{\boldsymbol{\zeta}}_j R_{\alpha_j^*} \in \mathbb{R}^2$  where  $\bar{\mathbf{f}} = \frac{1}{k} \sum_{\ell=1}^k \mathbf{f}_\ell \in \mathbb{R}^2$ ,  $\bar{\boldsymbol{\zeta}}_j = \frac{1}{k} \sum_{\ell=1}^k [\boldsymbol{\zeta}_j]_\ell \in \mathbb{R}^2$  and  $\bar{\mathbf{Y}} = \frac{1}{J} \sum_{j=1}^J \bar{\mathbf{Y}}_j = \bar{\mathbf{f}} (\overline{e^{a^*} R_{\alpha^*}}) + \bar{\mathbf{b}}^* +$

$\frac{1}{J} \sum_{j=1}^J e^{a_j^*} \bar{\zeta}_j R_{\alpha_j^*} \in \mathbb{R}^2$ . Thence, we have

$$\frac{1}{J} \|\mathbf{b}_{\Theta_0}^* - \hat{\mathbf{b}}^\lambda\|_{\mathbb{R}^{2J}}^2 \leq \frac{1}{J} \sum_{j=1}^J \|\bar{\mathbf{f}}(e^{a_j^*} R_{\alpha_j^*} - e^{\hat{a}_j^\lambda} \overline{(e^{a^*} R_{\alpha^*})} \overline{(e^{\hat{a}^\lambda} R_{\hat{\alpha}^\lambda})}^{-1} R_{\hat{\alpha}_j^\lambda})\|_{\mathbb{R}^2}^2 \quad (\text{A.11})$$

$$+ \|\bar{\mathbf{b}}^*(e^{a_j^*} \overline{(e^{a^*} R_{\alpha^*})}^{-1} R_{\alpha_j^*} - e^{\hat{a}_j^\lambda} \overline{(e^{\hat{a}^\lambda} R_{\hat{\alpha}^\lambda})}^{-1} R_{\hat{\alpha}_j^\lambda})\|_{\mathbb{R}^2}^2 \quad (\text{A.12})$$

$$+ \|e^{a_j^*} \bar{\zeta}_j R_{\alpha_j^*}\|_{\mathbb{R}^2}^2 + \|e^{\hat{a}_j^\lambda} \overline{(e^{a^*} \zeta R_{\alpha^*})} \overline{(e^{\hat{a}^\lambda} R_{\hat{\alpha}^\lambda})}^{-1} R_{\hat{\alpha}_j^\lambda}\|_{\mathbb{R}^2}^2 \quad (\text{A.13})$$

where  $\overline{(e^{a^*} \zeta R_{\alpha^*})} = \frac{1}{J} \sum_{j=1}^J e^{a_j^*} \bar{\zeta}_j R_{\alpha_j^*} \in \mathbb{R}^2$ . The rest of the proof is devoted to the control of the terms (A.11), (A.12) and (A.13).

In this section, we denote by  $\|\cdot\|_{op}$  the operator norm of a  $2 \times 2$  matrix, *i.e.*  $\|A\|_{op} = |\gamma_{\max}(A)|$  where  $\gamma_{\max}(A)$  denotes the largest eigenvalue of a matrix  $A \in \mathbb{R}^{2 \times 2}$ . Note by the way that the eigenvalues of the matrix  $\frac{1}{J} \sum_{j=1}^J e^{a_j} R_{\alpha_j} \in \mathbb{R}^{2 \times 2}$  are  $\frac{1}{J} \sum_{j=1}^J e^{a_j} (\cos(\alpha_j) \pm i \sin(\alpha_j))$  for any  $J$  and  $(a_1, \dots, a_J, \alpha_1, \dots, \alpha_J) \in [-A, A]^J \times [-\mathcal{A}, \mathcal{A}]^J$ . It yields  $\|e^{-a_j} R_{-\alpha_j}\|_{op} \leq e^A$  and  $\|\overline{(e^{\mathcal{A}} R_{\mathcal{A}})}^{-1}\|_{op} \leq e^A (\cos(\mathcal{A}) - \sin(\mathcal{A}))^{-1}$  which is a positive real number since by hypothesis we have  $\mathcal{A} < \frac{\pi}{4}$ . We are now able to derive an upper bound for (A.11),

$$\begin{aligned} (\text{A.11}) &\leq \|\bar{\mathbf{f}}\|_{\mathbb{R}^2}^2 \frac{1}{J} \sum_{j=1}^J \|e^{a_j^*} R_{\alpha_j^*} - e^{\hat{a}_j^\lambda} \overline{(e^{a^*} R_{\alpha^*})} \overline{(e^{\hat{a}^\lambda} R_{\hat{\alpha}^\lambda})}^{-1} R_{\hat{\alpha}_j^\lambda}\|_{op}^2 \\ &\leq e^{2A} \|\bar{\mathbf{f}}\|_{\mathbb{R}^2}^2 \frac{1}{J} \sum_{j=1}^J \|e^{a_j^*} R_{\alpha_j^*} (e^{-\hat{a}_j^\lambda} R_{-\hat{\alpha}_j^\lambda}) - \overline{(e^{a^*} R_{\alpha^*})} \overline{(e^{\hat{a}^\lambda} R_{\hat{\alpha}^\lambda})}^{-1}\|_{op}^2 \\ &\leq 2e^{2A} \|\bar{\mathbf{f}}\|_{\mathbb{R}^2}^2 \frac{1}{J} \sum_{j=1}^J \|e^{-\hat{a}_j^\lambda} R_{-\hat{\alpha}_j^\lambda}\|_{op}^2 \|e^{a_j^*} R_{\alpha_j^*} - e^{\hat{a}_j^\lambda} R_{\hat{\alpha}_j^\lambda}\|_{op}^2 \\ &\quad + 2e^{2A} \|\bar{\mathbf{f}}\|_{\mathbb{R}^2}^2 \|\overline{(e^{\hat{a}^\lambda} R_{\hat{\alpha}^\lambda})}^{-1}\|_{op}^2 \|\overline{(e^{a^*} R_{\alpha^*})} - \overline{(e^{\hat{a}^\lambda} R_{\hat{\alpha}^\lambda})}\|_{op}^2 \end{aligned}$$

It is now easy to show that there exists a constant  $C(A, \mathcal{A}, \mathbf{f}) > 0$  independent of  $k$  and  $J$  such that

$$(\text{A.11}) \leq C(A, \mathcal{A}, \mathbf{f}) \frac{1}{J} \sum_{j=1}^J \|(a_j^* - \hat{a}_j^\lambda, \alpha_j^* - \hat{\alpha}_j^\lambda)\|_{\mathbb{R}^2}^2. \quad (\text{A.14})$$

The term (A.12) is very similar to the term (A.11) and we have

$$(\text{A.12}) \leq C(A, \mathcal{A}, B) \frac{1}{J} \sum_{j=1}^J \|(a_j^* - \hat{a}_j^\lambda, \alpha_j^* - \hat{\alpha}_j^\lambda)\|_{\mathbb{R}^2}^2, \quad (\text{A.15})$$

for some constant  $C(A, \mathcal{A}, B) > 0$  independent of  $k$  and  $J$ . By using formula (A.14), (A.15) and Proposition A.1 together, there exists a constant  $C(L, s, A, \mathcal{A}, B, \mathbf{f})$  independent of  $k$  and  $J$  such that for all  $x > 0$ ,

$$\mathbb{P}\left((\text{A.11}) + (\text{A.12}) \geq C(L, s, A, \mathcal{A}, B, \mathbf{f}) (F(k^{-\frac{2s}{2s+1}}) + F(V_1(k, J, x)))\right) \leq 4e^{-x}, \quad (\text{A.16})$$

where  $V_1(k, J, x)$  is defined in statement of proposition A.1 and  $F(u) = u + \sqrt{u}$  for  $u \geq 0$ .

The term (A.13) can be bounded in probability as follows. First of all, remark that there exists a constant  $C(A, \mathcal{A}) > 0$  such that

$$(\text{A.13}) \leq C(A, \mathcal{A}) \frac{1}{J} \sum_{j=1}^J \|\bar{\zeta}_j\|_{\mathbb{R}^2}^2.$$

Then, for all  $j = 1, \dots, J$ , the random variable  $\bar{\zeta}_j = \frac{1}{k} \sum_{\ell=1}^k ([\zeta_j^{(1)}]_\ell, [\zeta_j^{(2)}]_\ell) \in \mathbb{R}^2$  can be written

$$(\bar{\zeta}_j^1, \bar{\zeta}_j^2)' = \frac{1}{k} \begin{pmatrix} \mathbb{1}'_k & 0'_k \\ 0'_k & \mathbb{1}'_k \end{pmatrix} ([\zeta_j^{(1)}]_1, \dots, [\zeta_j^{(1)}]_k, [\zeta_j^{(2)}]_1, \dots, [\zeta_j^{(2)}]_k)'$$

where  $0_k$  is a column vector of  $k$  zeros. The random vector  $(\bar{\zeta}_j^1, \bar{\zeta}_j^2)'$  is a two dimensional centered Gaussian vector of variance  $\mathbf{V} = \frac{1}{k^2} \begin{pmatrix} \mathbb{1}'_k & 0'_k \\ 0'_k & \mathbb{1}'_k \end{pmatrix} \Sigma \begin{pmatrix} \mathbb{1}_k & 0_k \\ 0_k & \mathbb{1}_k \end{pmatrix}$ . The  $2 \times 2$  matrix  $\mathbf{V}$  is of trace less or equal to  $\frac{2}{k} \gamma_{\max}(k)$ . Hence, the random variable  $\frac{1}{J} \sum_{j=1}^J \|\bar{\zeta}_j\|_{\mathbb{R}^2}^2 = \frac{1}{J} \sum_{j=1}^J (\bar{\zeta}_j^{(1)}, \bar{\zeta}_j^{(2)}) (\bar{\zeta}_j^{(1)}, \bar{\zeta}_j^{(2)})'$  has the same probability distribution as  $\frac{1}{J} \xi' [Id_J \otimes \mathbf{V}] \xi$  where  $\xi$  is a centered Gaussian vector of variance  $Id_{2J}$ . A standard concentration inequality of  $\chi^2$  distribution (see e.g. [Laurent & Massart, 2000] Lemma 1) is  $\mathbb{P} \left( \xi' [Id_J \otimes \mathbf{V}] \xi \geq J \frac{4}{k} \gamma_{\max}(k) + \frac{4}{k} \gamma_{\max}(k) \sqrt{2Jx} + x \frac{4}{k} \gamma_{\max}(k) \right) \leq e^{-x}$  for any  $x > 0$ . It yields that there exists a constant  $C(A, \mathcal{A}) > 0$  such that for all  $x > 0$

$$\mathbb{P} \left( (A.13) \geq C(A, \mathcal{A}) \frac{\gamma_{\max}(k)}{k} \left( 1 + \sqrt{2 \frac{x}{J}} + \frac{x}{J} \right) \right) \leq e^{-x}. \quad (A.17)$$

To end the proof remark that for all  $u \geq 0$  we have  $\mathbb{P}(\frac{1}{J} \|\mathbf{b}_{\Theta_0}^* - \hat{\mathbf{b}}^\lambda\|_{\mathbb{R}^{2J}}^2 \geq u) \leq \mathbb{P}((A.11) + (A.12) + (A.13) \geq u)$ . This together with (A.16) and (A.17) yield that there exists a constant  $C(L, s, A, \mathcal{A}, B, \mathbf{f}) > 0$  independent of  $k$  and  $J$  such that for all  $x > 0$  we have

$$\mathbb{P} \left( \frac{1}{J} \|\mathbf{b}_{\Theta_0}^* - \hat{\mathbf{b}}^\lambda\|_{\mathbb{R}^{2J}}^2 \geq C(L, s, A, \mathcal{A}, B, \mathbf{f}) (F(k^{-\frac{2s}{2s+1}}) + F(V_1(k, J, x)) + V_2(k, J, x)) \right) \leq 5e^{-x},$$

where  $V_2(k, J, x) = \frac{\gamma_{\max}(k)}{k} (1 + \sqrt{2 \frac{x}{J}} + \frac{x}{J})$  and the proof of Proposition A.3 is completed.  $\square$

## B Technical Lemma

**Lemma B.1.** Assume that Assumption 1 holds, i.e.  $f \in H_s(L)$  with  $s > \frac{3}{2}$  (see (1.3)) and  $\mathbf{f} = (f(\frac{\ell}{k}))_{\ell=1}^k \in \mathbb{R}^{k \times 2}$ . If  $\lambda(k) = k^{\frac{1}{2s+1}}$  then there exists a constant  $C(L, s)$  such that for all  $f \in H_s(L)$  we have

$$\frac{1}{k} \|\mathbf{A}^\lambda \mathbf{f} - \mathbf{f}\|_{\mathbb{R}^{k \times 2}}^2 \leq C(L, s) k^{-\frac{2s}{2s+1}},$$

where  $\mathbf{A}^\lambda$  is the projection matrix defined in (4.2).

*Proof.* Recall the notations introduced Sections 1.1 and 4.1 :  $c_m(f) = (c_m(f^{(1)}), c_m(f^{(2)})) \in \mathbb{C}^2$  is the  $m$ -th Fourier coefficient of  $f \in L^2([0, 1], \mathbb{R}^2)$  and  $c_m(\mathbf{f}) = (c_m^{(1)}(\mathbf{f}), c_m^{(2)}(\mathbf{f})) \in \mathbb{C}^2$  is the  $m$ -th discrete Fourier coefficient of  $\mathbf{f} \in \mathbb{R}^{k \times 2}$ . Thus, we have by Parseval's equality

$$\frac{1}{k} \left\| \mathbf{A}^\lambda \mathbf{f} - \mathbf{f} \right\|_{\mathbb{R}^{k \times 2}}^2 = \frac{1}{k} \left\| \left( \frac{1}{k} \sum_{|m| > \lambda} c_m(\mathbf{f}) e^{i2\pi m \frac{\ell}{k}} \right)_{\ell=1}^k \right\|_{\mathbb{R}^{k \times 2}}^2 = \frac{1}{k^2} \sum_{|m| > \lambda} \left\| c_m(\mathbf{f}) \right\|_{\mathbb{C}^2}^2.$$

where for all  $c = (c^{(1)}, c^{(2)}) \in \mathbb{C}^2$ ,  $\|c\|_{\mathbb{C}^2}^2 = |c^{(1)}|^2 + |c^{(2)}|^2$ . It yields

$$\frac{1}{k} \|\mathbf{A}^\lambda \mathbf{f} - \mathbf{f}\|_{\mathbb{R}^{k \times 2}}^2 \leq 2 \sum_{|m| > \lambda} \left\| \frac{1}{k} c_m(\mathbf{f}) - c_m(f) \right\|_{\mathbb{C}^2}^2 + 2 \sum_{|m| > \lambda} \|c_m(f)\|_{\mathbb{C}^2}^2.$$

Firstly, Lemma 1.10 in [Tsybakov, 2009] ensures that if  $s > \frac{1}{2}$  then  $|\frac{1}{k} c_m(\mathbf{f}^{(i)}) - c_m(f^{(i)})| \leq C(L, s) k^{-s + \frac{1}{2}}$  for any  $m \in \mathbb{N}$  and  $i = 1, 2$ . Secondly, equation (1.87) in [Tsybakov, 2009] ensure that if  $\lambda(k) = k^{\frac{1}{2s+1}}$  then  $\sum_{|m| > \lambda} |c_m(f^{(i)})|^2 \leq C(L, s) k^{-\frac{2s}{2s+1}}$ . Thence, there is a constant  $C(L, s)$  independent of  $k$  such that

$$\frac{1}{k} \|\mathbf{A}^\lambda \mathbf{f} - \mathbf{f}\|_{\mathbb{R}^{k \times 2}}^2 \leq C(L, s) (k^{2-2s} + k^{-\frac{2s}{2s+1}}).$$

If  $s \geq \frac{3}{2}$  then we have

$$\frac{1}{k} \|\mathbf{A}^\lambda \mathbf{f} - \mathbf{f}\|_{\mathbb{R}^{k \times 2}}^2 \leq C(L, s) k^{-\frac{2s}{2s+1}}. \quad \square$$

We now give a general expression of the gradient and of the Hessian of the criterion  $D$ . Recall that we have

$$D(\mathbf{a}, \boldsymbol{\alpha}, \mathbf{b}) = \frac{1}{J} \sum_{j=1}^J \left\| e^{-a_j} (e^{a_j^*} \mathbf{f} R_{\alpha_j^*} + \mathbb{1}_k \otimes (b_j^* - b_j)) R_{-\alpha_j} + \frac{1}{J} \sum_{j'=1}^J e^{-a_{j'}} (e^{a_{j'}^*} \mathbf{f} R_{\alpha_{j'}^*} + \mathbb{1}_k \otimes (b_{j'}^* - b_{j'})) R_{-\alpha_{j'}} \right\|_{\mathbb{R}^{k \times 2}}^2$$

where  $(\mathbf{a}, \boldsymbol{\alpha}, \mathbf{b}) = (a_1, \dots, a_J, \alpha_1, \dots, \alpha_J, b_1, \dots, b_J) \in \mathbb{R}^J \times [-\pi, \pi]^J \times \mathbb{R}^{2J}$ . To shorten the formulas below we note  $g_j = (g_j^{(1)}, g_j^{(2)}, g_j^{(3)}, g_j^{(4)}) = (a_j, \alpha_j, b_j)$ , that is  $g_j^{(1)} = a_j$ ,  $g_j^{(2)} = \alpha_j$ ,  $g_j^{(3)} = b_j^{(1)}$  and  $g_j^{(4)} = b_j^{(2)}$ . Let  $\mathbf{f}_{g_j} = g^{-1} \cdot g_j^* \cdot \mathbf{f} = e^{-a_j} (e^{a_j^*} \mathbf{f} R_{\alpha_j^*} + \mathbb{1}_k \otimes (b_j^* - b_j)) R_{-\alpha_j}$ , and for all  $j_1 = 1, \dots, J$  and  $p_1 = 1, \dots, 4$ ,

$$\partial_{g_{j_1}^{(p_1)}} D(\mathbf{a}, \boldsymbol{\alpha}, \mathbf{b}) = \frac{2}{Jk} \left\langle \partial_{g_{j_1}^{(p_1)}} \mathbf{f}_{g_{j_1}}, \mathbf{f}_{g_{j_1}} - \frac{1}{J} \sum_{j'=1}^J \mathbf{f}_{g_{j'}} \right\rangle_{\mathbb{R}^{k \times 2}}. \quad (\text{B.1})$$

The second order derivatives are

$$\partial_{g_{j_2}^{(p_2)}} \partial_{g_{j_1}^{(p_1)}} D(\mathbf{a}, \boldsymbol{\alpha}, \mathbf{b}) = -\frac{2}{J^2 k} \left\langle \partial_{g_{j_1}^{(p_1)}} \mathbf{f}_{g_{j_1}}, \partial_{g_{j_2}^{(p_2)}} \mathbf{f}_{g_{j_2}} \right\rangle_{\mathbb{R}^{k \times 2}}, \quad \text{if } j_1 \neq j_2, \quad (\text{B.2})$$

$$\begin{aligned} \partial_{g_{j_1}^{(p_2)}} \partial_{g_{j_1}^{(p_1)}} D(\mathbf{a}, \boldsymbol{\alpha}, \mathbf{b}) &= \frac{2}{Jk} \left\langle \partial_{g_{j_1}^{(p_2)}} \partial_{g_{j_1}^{(p_1)}} \mathbf{f}_{g_{j_1}}, \left( \mathbf{f}_{g_{j_1}} - \frac{1}{J} \sum_{j'=1}^J \mathbf{f}_{g_{j'}} \right) \right\rangle_{\mathbb{R}^{k \times 2}} \\ &\quad + \left( \frac{2}{Jk} - \frac{2}{J^2 k} \right) \left\langle \partial_{g_{j_1}^{(p_1)}} \mathbf{f}_{g_{j_1}}, \partial_{g_{j_1}^{(p_2)}} \mathbf{f}_{g_{j_1}} \right\rangle_{\mathbb{R}^{k \times 2}}. \end{aligned} \quad (\text{B.3})$$

The expressions of the gradient and the Hessian of  $D$  simplify on the set  $(\mathbf{a}^*, \boldsymbol{\alpha}^*, \mathbf{b}^*) * \mathcal{G}$ , see Lemma 3.1. For any  $g_0 = (a_0, \alpha_0, b_0) \in \mathbb{R} \times [-\pi, \pi] \times \mathbb{R}^2$ , we have  $\mathbf{f}_{g_{j_1}^* \cdot g_0} - \frac{1}{J} \sum_{j'=1}^J \mathbf{f}_{g_{j'}^* \cdot g_0} = e^{-a_0} (\mathbf{f} - \mathbb{1}_k \otimes b_0) R_{-\alpha_0} - \frac{1}{J} \sum_{j'=1}^J e^{-a_0} (\mathbf{f} - \mathbb{1}_k \otimes b_0) R_{-\alpha_0} = 0$ . It yields that for all  $(a_0, \alpha_0, b_0) \in \mathbb{R} \times [-\pi, \pi] \times \mathbb{R}^2$  we have  $\nabla D((\mathbf{a}^*, \boldsymbol{\alpha}^*, \mathbf{b}^*) * (a_0, \alpha_0, b_0)) = 0$ , and

$$\partial_{g_{j_2}^{(p_2)}} \partial_{g_{j_1}^{(p_1)}} D((\mathbf{a}^*, \boldsymbol{\alpha}^*, \mathbf{b}^*) * (a_0, \alpha_0, b_0)) = \begin{cases} -\frac{2}{J^2 k} \left\langle \partial_{g_{j_1}^{(p_1)}} \mathbf{f}_{g_{j_1}^* \cdot g_0}, \partial_{g_{j_2}^{(p_2)}} \mathbf{f}_{g_{j_2}^* \cdot g_0} \right\rangle_{\mathbb{R}^{k \times 2}}, & \text{if } j_1 \neq j_2, \\ \left( \frac{2}{Jk} - \frac{2}{J^2 k} \right) \left\langle \partial_{g_{j_1}^{(p_1)}} \mathbf{f}_{g_{j_1}^* \cdot g_0}, \partial_{g_{j_2}^{(p_2)}} \mathbf{f}_{g_{j_1}^* \cdot g_0} \right\rangle_{\mathbb{R}^{k \times 2}}. \end{cases} \quad (\text{B.4})$$

**Lemma B.2.** *The smallest eigenvalue of  $\nabla^2 D_0(\mathbf{a}_{\Theta_0}^*, \boldsymbol{\alpha}_{\Theta_0}^*)$  restricted to the subset  $\Theta_0$  is greater than  $e^{-2A} \frac{2}{Jk} \|\mathbf{f}_0\|_{\mathbb{R}^{k \times 2}}^2$ .*

*Proof.* In this proof,  $\mathbf{h} = (\mathbf{a}, \boldsymbol{\alpha}) \in \mathbb{R}^J \times [-\pi, \pi]^J$  and  $\mathbf{f}_{h_j}^0 = e^{a_j^* - a_j} \mathbf{f} R_{\alpha_j^* - \alpha_j}$ . We have  $\partial_{a_{j_1}} \mathbf{f}_{h_{j_1}}^0 = -\mathbf{f}_{h_{j_1}}^0$  and  $\partial_{\alpha_{j_1}} \mathbf{f}_{h_{j_1}}^0 = \mathbf{f}_{h_{j_1}}^0 R_{-\frac{\pi}{2}}$  for all  $j_1 = 1, \dots, J$ . By using Formulas (B.4), the Hessian of  $D_0$  at the point  $(\mathbf{a}_{\Theta_0}^*, \boldsymbol{\alpha}_{\Theta_0}^*) = (\mathbf{a}^* - \bar{\mathbf{a}}^*, \boldsymbol{\alpha}^* - \bar{\boldsymbol{\alpha}}^*) \in \Theta_0^{\mathbf{a}, \boldsymbol{\alpha}}$  is given by,

$$\partial_{\alpha_{j_1}} \partial_{\alpha_{j_2}} D_0(\mathbf{a}^* - \bar{\mathbf{a}}^*, \boldsymbol{\alpha}^* - \bar{\boldsymbol{\alpha}}^*) = \partial_{a_{j_1}} \partial_{a_{j_2}} D_0(\mathbf{a}^* - \bar{\mathbf{a}}^*, \boldsymbol{\alpha}^* - \bar{\boldsymbol{\alpha}}^*) = \begin{cases} -\frac{2}{J^2 k} \|e^{\bar{\mathbf{a}}^*} \mathbf{f}_0\|_{\mathbb{R}^{k \times 2}}^2, & \text{if } j_1 \neq j_2, \\ \left( \frac{2}{Jk} - \frac{2}{J^2 k} \right) \|e^{\bar{\mathbf{a}}^*} \mathbf{f}_0\|_{\mathbb{R}^{k \times 2}}^2 \end{cases}$$

and the second order cross derivatives are 0. The Hessian of  $D_0$  at  $(\mathbf{a}^* - \bar{\mathbf{a}}^*, \boldsymbol{\alpha}^* - \bar{\boldsymbol{\alpha}}^*)$  can be written,

$$\nabla^2 D_0(\mathbf{a}^* - \bar{\mathbf{a}}^*, \boldsymbol{\alpha}^* - \bar{\boldsymbol{\alpha}}^*) = \frac{2}{J^2 k} \|e^{\bar{\mathbf{a}}^*} \mathbf{f}_0\|_{\mathbb{R}^{k \times 2}}^2 \begin{pmatrix} JId_J - \mathbb{1}_{J \times J} & 0 \\ 0 & JId_J - \mathbb{1}_{J \times J} \end{pmatrix}, \quad (\text{B.5})$$

where  $Id_J$  is the identity in  $\mathbb{R}^J$  and  $\mathbb{1}_J$  is the  $J \times J$  matrix of ones. The eigenvalues of  $JId_J - \mathbb{1}_{J \times J}$  are 0 with eigenvector  $\mathbb{1}_J$  and  $J$  with eigenspace  $\mathbb{1}_J^\perp$ . It yields that on  $\Theta_0^{\mathbf{a}, \boldsymbol{\alpha}}$  the smallest eigenvalue of  $\nabla^2 D_0(\mathbf{a}^* - \bar{\mathbf{a}}^*, \boldsymbol{\alpha}^* - \bar{\boldsymbol{\alpha}}^*)$  is  $e^{2\bar{\mathbf{a}}^*} \frac{2}{Jk} \|\mathbf{f}_0\|_{\mathbb{R}^{k \times 2}}^2$ . To finish the proof, remark that  $\bar{\mathbf{a}}^* \leq A$ .  $\square$

**Lemma B.3.** Let  $\delta = \max\{A, \mathcal{A}\}$ . For all  $\mathbf{c} = (a_1, \dots, a_J, \alpha_1, \dots, \alpha_J) \in \Theta_0^{\mathbf{a}, \alpha}$  and  $(\mathbf{a}, \alpha) \in \mathbb{R}^{2J}$  we have

$$|[\nabla^3 D_0(\mathbf{c})](\mathbf{a}, \alpha)| \leq 40\delta e^{2A} \|\mathbf{f}_0\|_{\mathbb{R}^{k \times 2}}^2 \frac{1}{Jk} \|(\mathbf{a}, \alpha)\|_{\mathbb{R}^{2J}}^2.$$

*Proof.* In this proof,  $\mathbf{h} = (\mathbf{a}, \alpha) \in \mathbb{R}^J \times [-\pi, \pi]^J$ , that is  $h_j^{(1)} = a_j$  is the parameter of scaling and  $h_j^{(2)} = \alpha_j$  is the parameter of rotation. We have  $\mathbf{f}_{h_j}^0 = e^{a_j^* - a_j} \mathbf{f}_0 R_{\alpha_j^* - \alpha_j}$ . Then, from equations (B.2) and (B.3), it follows that for all  $j_1, j_2, j_3 = 1, \dots, J$  and  $p_1, p_2, p_3 = 1, \dots, 2$ ,

$$\begin{aligned} \partial_{h_{j_3}^{(p_3)}} \partial_{h_{j_2}^{(p_2)}} \partial_{h_{j_1}^{(p_1)}} D_0(\mathbf{a}, \alpha) &= 0, \quad \text{if } j_1 \neq j_2 \text{ and } j_2 \neq j_3 \text{ and } j_1 \neq j_3, \\ \partial_{h_{j_2}^{(p_2)}} \partial_{h_{j_1}^{(p_1)}} \partial_{h_{j_1}^{(p_1)}} D_0(\mathbf{a}, \alpha) &= -\frac{2}{J^2 k} \left\langle \partial_{h_{j_1}^{(p_2)}} \partial_{h_{j_1}^{(p_1)}} \mathbf{f}_{h_{j_1}}^0, \partial_{h_{j_2}^{(p_3)}} \mathbf{f}_{h_{j_2}}^0 \right\rangle_{\mathbb{R}^{k \times 2}}, \quad \text{if } j_1 \neq j_2, \\ \partial_{h_{j_1}^{(p_3)}} \partial_{h_{j_1}^{(p_2)}} \partial_{h_{j_1}^{(p_1)}} D_0(\mathbf{a}, \alpha) &= \frac{2}{Jk} \left\langle \partial_{h_{j_1}^{(p_3)}} \partial_{h_{j_1}^{(p_2)}} \partial_{h_{j_1}^{(p_1)}} \mathbf{f}_{h_{j_1}}^0, \left( \mathbf{f}_{h_{j_1}}^0 - \frac{1}{J} \sum_{j'=1}^J \mathbf{f}_{h_{j'}}^0 \right) \right\rangle_{\mathbb{R}^{k \times 2}} \\ &\quad + \left( \frac{2}{Jk} - \frac{2}{J^2 k} \right) \left( \left\langle \partial_{h_{j_1}^{(p_2)}} \partial_{h_{j_1}^{(p_1)}} \mathbf{f}_{h_{j_1}}^0, \partial_{h_{j_1}^{(p_3)}} \mathbf{f}_{h_{j_1}}^0 \right\rangle_{\mathbb{R}^{k \times 2}} + \left\langle \partial_{h_{j_1}^{(p_3)}} \partial_{h_{j_1}^{(p_1)}} \mathbf{f}_{h_{j_1}}^0, \right. \right. \\ &\quad \left. \left. \partial_{h_{j_1}^{(p_2)}} \mathbf{f}_{h_{j_1}}^0 \right\rangle_{\mathbb{R}^{k \times 2}} + \left\langle \partial_{h_{j_1}^{(p_3)}} \partial_{h_{j_1}^{(p_2)}} \mathbf{f}_{h_{j_1}}^0, \partial_{h_{j_1}^{(p_1)}} \mathbf{f}_{h_{j_1}}^0 \right\rangle_{\mathbb{R}^{k \times 2}} \right) \end{aligned}$$

By Cauchy-Schwarz inequality we have,

$$\left| \left\langle \partial_{h_{j_1}^{(p_2)}} \partial_{h_{j_1}^{(p_1)}} \mathbf{f}_{h_{j_1}}^0, \partial_{h_{j_2}^{(p_3)}} \mathbf{f}_{h_{j_2}}^0 \right\rangle_{\mathbb{R}^{k \times 2}} \right| \leq \|\partial_{h_{j_1}^{(p_2)}} \partial_{h_{j_1}^{(p_1)}} \mathbf{f}_{h_{j_1}}^0\|_{\mathbb{R}^{k \times 2}} \|\partial_{h_{j_2}^{(p_3)}} \mathbf{f}_{h_{j_2}}^0\|_{\mathbb{R}^{k \times 2}} \leq e^{2A} \|\mathbf{f}_0\|_{\mathbb{R}^{k \times 2}}^2. \quad (\text{B.6})$$

and

$$\begin{aligned} &\left| \left\langle \partial_{h_{j_1}^{(p_3)}} \partial_{h_{j_1}^{(p_2)}} \partial_{h_{j_1}^{(p_1)}} \mathbf{f}_{h_{j_1}}^0, \left( \mathbf{f}_{h_{j_1}}^0 - \frac{1}{J} \sum_{j'=1}^J \mathbf{f}_{h_{j'}}^0 \right) \right\rangle_{\mathbb{R}^{k \times 2}} \right| \\ &\leq \left\| \partial_{h_{j_1}^{(p_3)}} \partial_{h_{j_1}^{(p_2)}} \partial_{h_{j_1}^{(p_1)}} \mathbf{f}_{h_{j_1}}^0 \right\|_{\mathbb{R}^{k \times 2}} \left\| \mathbf{f}_{h_{j_1}}^0 - \frac{1}{J} \sum_{j'=1}^J \mathbf{f}_{h_{j'}}^0 \right\|_{\mathbb{R}^{k \times 2}} \leq 2e^{2A} \|\mathbf{f}_0\|_{\mathbb{R}^{k \times 2}}^2 \quad (\text{B.7}) \end{aligned}$$

For  $\kappa = (\kappa_1, \dots, \kappa_{2J}) \in \mathbb{N}^{2J}$ , denote by  $|\kappa| = \kappa_1 + \dots + \kappa_{2J}$  and

$$(\partial_{\mathbf{h}})^\kappa = (\partial_{a_1})^{\kappa_1} (\partial_{\alpha_1})^{\kappa_2} \dots (\partial_{a_J})^{\kappa_{2J-1}} (\partial_{\alpha_J})^{\kappa_{2J}}.$$

Then, the differential of order 3 of  $D_0$  at  $\mathbf{c} \in \Theta_0^{\mathbf{a}, \alpha}$  applied at  $(\mathbf{a}, \alpha) \in \mathbb{R}^{2J}$  writes as

$$[\nabla^3 D_0(\mathbf{c})](\mathbf{a}, \alpha) = \sum_{|\kappa|=3} (\partial_{\mathbf{h}})^\kappa D_0(\mathbf{c}) \mathbf{h}^\kappa$$

where  $\mathbf{h}^\kappa = a_1^{\kappa_1} \alpha_1^{\kappa_2} \dots a_J^{\kappa_{2J-1}} \alpha_J^{\kappa_{2J}}$ . This formula together with equations (B.6) and (B.7) give,

$$\begin{aligned}
\left| [\nabla^3 D_0(\mathbf{c})](\mathbf{a}, \boldsymbol{\alpha}) \right| &= \left| \sum_{p_1, p_2, p_3=1}^2 \sum_{j_1=1}^J \partial_{h_{j_1}^{(p_3)}} \partial_{h_{j_1}^{(p_2)}} \partial_{h_{j_1}^{(p_1)}} D_0(\mathbf{c}) h_{j_1}^{(p_1)} h_{j_1}^{(p_2)} h_{j_1}^{(p_3)} \right. \\
&\quad \left. + 3 \sum_{j_1 \neq j_2=1}^J \partial_{h_{j_2}^{(p_3)}} \partial_{h_{j_1}^{(p_2)}} \partial_{h_{j_1}^{(p_1)}} D_0(\mathbf{c}) h_{j_1}^{(p_1)} h_{j_1}^{(p_2)} h_{j_2}^{(p_3)} \right| \\
&\leq 2e^{2A} \frac{1}{k} \|\mathbf{f}_0\|_{\mathbb{R}^{k \times 2}}^2 \sum_{p_1, p_2, p_3=1}^2 \left( \frac{4}{J} \sum_{j_1=1}^J |h_{j_1}^{(p_1)} h_{j_1}^{(p_2)} h_{j_1}^{(p_3)}| + \frac{6}{J^2} \sum_{j_1 \neq j_2=1}^J |h_{j_1}^{(p_1)} h_{j_1}^{(p_2)} h_{j_2}^{(p_3)}| \right) \\
&\leq 2\delta e^{2A} \frac{1}{k} \|\mathbf{f}_0\|_{\mathbb{R}^{k \times 2}}^2 \sum_{p_1, p_2=1}^2 \left( \frac{4}{J} \sum_{j_1=1}^J |h_{j_1}^{(p_1)} h_{j_1}^{(p_2)}| + \frac{6(J-1)}{J^2} \sum_{j_1=1}^J |h_{j_1}^{(p_1)} h_{j_1}^{(p_2)}| \right) \\
&\leq 20\delta e^{2A} \frac{1}{k} \|\mathbf{f}_0\|_{\mathbb{R}^{k \times 2}}^2 \frac{1}{J} \sum_{j=1}^J \sum_{p_1, p_2=1}^2 |h_j^{(p_1)} h_j^{(p_2)}| \\
&= 20\delta e^{2A} \frac{1}{k} \|\mathbf{f}_0\|_{\mathbb{R}^{k \times 2}}^2 \frac{1}{J} \sum_{j=1}^J \left( \sum_{p_1=1}^2 |h_j^{(p_1)}| \right)^2 \leq 40\delta e^{2A} \frac{1}{k} \|\mathbf{f}_0\|_{\mathbb{R}^{k \times 2}}^2 \frac{1}{J} \sum_{j=1}^J \sum_{p_1=1}^2 |h_j^{(p_1)}|^2.
\end{aligned}$$

And the claim is now proved.  $\square$

**Lemma B.4.** For all  $\mathbf{f} \in \mathbb{R}^{k \times 2}$  and  $(a_1, \alpha_1, b_1), (a_2, \alpha_2, b_2) \in [-A, A] \times [-\mathcal{A}, \mathcal{A}] \times [-B, B]^2$ , let  $g_i \cdot \mathbf{f} = e^{a_i} \mathbf{f} R_{\alpha_i} + \mathbb{1}_k \otimes b_i$ ,  $i = 1, 2$ . Then, we have

$$\frac{1}{k} \|g_1 \cdot \mathbf{f} - g_2 \cdot \mathbf{f}\|_{\mathbb{R}^{k \times 2}}^2 \leq C(A, \mathbf{f}) \|(a_1, \alpha_1, b_1) - (a_2, \alpha_2, b_2)\|_{\mathbb{R}^4}^2,$$

where  $C(A, \mathbf{f}) = 2 \max\{4e^{4A} \frac{1}{k} \|\mathbf{f}\|_{\mathbb{R}^{k \times 2}}^2, 1\}$ .

*Proof.* We have

$$\frac{1}{k} \|g_1 \cdot \mathbf{f} - g_2 \cdot \mathbf{f}\|_{\mathbb{R}^{k \times 2}}^2 \leq 2e^{2A} \frac{1}{k} \|e^{a_1 - a_2} \mathbf{f} R_{\alpha_1 - \alpha_2} - \mathbf{f}\|_{\mathbb{R}^{k \times 2}}^2 + 2 \frac{1}{k} \|\mathbb{1}_k \otimes (b_1 - b_2)\|_{\mathbb{R}^{k \times 2}}^2 \quad (\text{B.8})$$

Let now  $F(a, \alpha) = \frac{1}{\sqrt{k}} \|e^a \mathbf{f} R_\alpha - \mathbf{f}\|_{\mathbb{R}^{k \times 2}}$ . We have  $|\partial_a F(a, \alpha)| = \frac{1}{\sqrt{k} \|e^a \mathbf{f} R_\alpha - \mathbf{f}\|_{\mathbb{R}^{k \times 2}}} |\langle e^a \mathbf{f} R_\alpha, e^a \mathbf{f} R_\alpha - \mathbf{f} \rangle_{\mathbb{R}^{k \times 2}}| \leq \frac{e^A}{\sqrt{k}} \|\mathbf{f}\|_{\mathbb{R}^{k \times 2}}$  and  $|\partial_\alpha F(a, \alpha)| = \frac{1}{\sqrt{k} \|e^a \mathbf{f} R_\alpha - \mathbf{f}\|_{\mathbb{R}^{k \times 2}}} |\langle e^a \mathbf{f} R_{\alpha + \frac{\pi}{2}}, e^a \mathbf{f} R_\alpha - \mathbf{f} \rangle_{\mathbb{R}^{k \times 2}}| \leq \frac{e^A}{\sqrt{k}} \|\mathbf{f}\|_{\mathbb{R}^{k \times 2}}$ . The Euclidean norm of the gradient of  $F$  satisfies

$$\|\nabla F(a, \alpha)\|_{\mathbb{R}^2} = \sqrt{|\partial_a F(a, \alpha)|^2 + |\partial_\alpha F(a, \alpha)|^2} \leq \sqrt{2} e^A \frac{1}{\sqrt{k}} \|\mathbf{f}\|_{\mathbb{R}^{k \times 2}}.$$

Since we have  $|F(a, \alpha)| = |F(a, \alpha) - F(0, 0)| \leq \sqrt{2} e^A \frac{1}{\sqrt{k}} \|\mathbf{f}\|_{\mathbb{R}^{k \times 2}} \|(a, \alpha)\|_{\mathbb{R}^2}$ , equation (B.8) yields

$$\frac{1}{k} \|g_1 \cdot \mathbf{f} - g_2 \cdot \mathbf{f}\|_{\mathbb{R}^{k \times 2}}^2 \leq 2 \max\left\{4e^{4A} \frac{1}{k} \|\mathbf{f}\|_{\mathbb{R}^{k \times 2}}^2, 1\right\} (|a_1 - a_2|^2 + |\alpha_1 - \alpha_2|^2 + |b_1^{(1)} - b_2^{(1)}|^2 + |b_1^{(2)} - b_2^{(2)}|^2),$$

which concludes the proof.  $\square$

## References

[Bhattacharya & Patrangenaru, 2003] Bhattacharya, R. & Patrangenaru, V. (2003). Large sample theory of intrinsic and extrinsic sample means on manifolds. i. *The Annals of Statistics*, 31(1), pp. 1–29.

- [Bigot & Charlier, 2011] Bigot, J. & Charlier, B. (2011). On the consistency of fréchet means in deformable models for curve and image analysis. *Electronic Journal of Statistics*, To appear.
- [Dryden & Mardia, 1998] Dryden, I. L. & Mardia, K. V. (1998). *Statistical shape analysis*. Wiley Series in Probability and Statistics: Probability and Statistics. Chichester: John Wiley & Sons Ltd.
- [Goodall, 1991] Goodall, C. (1991). Procrustes methods in the statistical analysis of shape. *J. Roy. Statist. Soc. Ser. B*, 53(2), 285–339.
- [Gray, 2006] Gray, M. (2006). *Toeplitz and circulant matrices: a review*. Now Publishers Inc.
- [Horn & Johnson, 1990] Horn, R. & Johnson, C. (1990). *Matrix analysis*. Cambridge: Cambridge University Press.
- [Huckemann, 2011] Huckemann, S. (2011). Inference on 3d procrustes means: Tree bole growth, rank deficient diffusion tensors and perturbation models. *Scand. J. Statist.*, To appear.
- [Kendall, 1984] Kendall, D. G. (1984). Shape manifolds, Procrustean metrics, and complex projective spaces. *Bull. London Math. Soc.*, 16(2), 81–121.
- [Kendall et al., 1999] Kendall, D. G., Barden, D., Carne, T. K., & Le, H. (1999). *Shape and shape theory*. Wiley Series in Probability and Statistics. Chichester: John Wiley & Sons Ltd.
- [Kent, 1992] Kent, J. T. (1992). New directions in shape analysis. In *The art of statistical science*, Wiley Ser. Probab. Math. Statist. Probab. Math. Statist. (pp. 115–127). Chichester: Wiley.
- [Kent & Mardia, 1997] Kent, J. T. & Mardia, K. V. (1997). Consistency of Procrustes estimators. *J. Roy. Statist. Soc. Ser. B*, 59(1), 281–290.
- [Laurent & Massart, 2000] Laurent, B. & Massart, P. (2000). Adaptive estimation of a quadratic functional by model selection. *Ann. Stat.*, 28(5), 1302–1338.
- [Le, 1998] Le, H. (1998). On the consistency of procrustean mean shapes. *Adv. in Appl. Probab.*, 30(1), 53–63.
- [Lele, 1993] Lele, S. (1993). Euclidean distance matrix analysis (EDMA): estimation of mean form and mean form difference. *Math. Geol.*, 25(5), 573–602.
- [Massart, 2007] Massart, P. (2007). *Concentration inequalities and model selection*, volume 1896 of *Lecture Notes in Mathematics*. Berlin: Springer.
- [Tsybakov, 2009] Tsybakov, A. (2009). *Introduction to nonparametric estimation*. Springer Series in Statistics. New York, NY: Springer. xii, 214 p.
- [van der Vaart, 1998] van der Vaart, A. W. (1998). *Asymptotic statistics*, volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge: Cambridge University Press.